



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 95836

TO: Cynthia Collins
Location: cm1/9a12/9e12
Art Unit: 1638
Thursday, June 05, 2003

Case Serial Number: 854122

From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203

edward.hart@uspto.gov

Search Notes

Examiner Collins,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or *contact:*

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



THIS PAGE BLANK (USPTO)

95836

STIC-Biotech/ChemLib

From: Collins, Cynthia
Sent: Wednesday, June 04, 2003 6:45 PM
To: STIC-Biotech/ChemLib
Subject: sequence search request SN 09/854122

RECEIVED

JUN - 5 2003

Please search prior art for SN 09/854122:

(STIC)

1) SEQ ID NO:1

Thank You,

Cynthia Collins
Art Unit 1638
CM1, 9A12 (office) or 9E12 (mailbox)
(703) 605-1210

Edward Han
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

TYPE OF SEARCH:

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 6/5/03
Date Completed: 6/5/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:16:42 ; Search time 70 Seconds
(without alignments)
441.631 Million cell updates/sec

Title: US-09-854-122-1
Perfect score: 320
Sequence: 1 TVPKSGTXXXXXXX.....XXXXXXXXXGDKWXXFT 232

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	16.9	295	18	AAW23657 E6AP-binding prote
2	54	16.9	295	21	AAV67294 Human STP2 (phenol
3	54	16.9	1305	22	ABG27554 Novel human diagno
4	52	16.2	129	21	AAB45369 Gene 35 human secr
5	52	16.2	129	21	AAB45370 Human secreted pro
6	52	16.2	141	22	AA675324 Human colon cancer
7	52	16.2	294	19	AAW40498 Human EST protein.
8	52	16.2	294	19	AAW44247 Human oestrogen su
9	51	15.9	13	19	AAW40502 Human EST protein
10	51	15.9	13	23	AAU78272 SUL7n PAPS 3' phos

11	51	15.9	23	AAU12052	Sulfotransferase R
12	50	15.6	104	21	AAG03995 Human secreted pro
13	50	15.6	283	23	ABB81794 Human sulfotransfe
14	50	15.6	304	22	AAE12208 Human ST drug-meta
15	50	15.6	304	22	AAE12209 Human ST drug-meta
16	50	15.6	304	22	AAE12210 Human ST drug-meta
17	50	15.6	304	22	AAE05178 Human drug metabol
18	50	15.6	304	23	ABB81792 Human sulfotransfe
19	50	15.6	304	23	ABB81793 Human sulfotransfe
20	50	15.6	304	23	AAE08043 Orphan G protein-c
21	50	15.6	305	23	AAE08043 Human POLY13 prote
22	47	14.7	214	22	AAU07765 Human novel transf
23	47	14.7	265	22	AAU07760 Human novel transf
24	47	14.7	303	22	AAU07758 Arabidopsis thalia
25	46	14.4	186	21	AAG21658 Arabidopsis thalia
26	46	14.4	201	21	AAG16880 Arabidopsis thalia
27	46	14.4	327	21	AAG16879 Arabidopsis thalia
28	46	14.4	331	21	AAG21657 Arabidopsis thalia
29	46	14.4	333	23	ABB91753 Arabidopsis thalia
30	46	14.4	337	21	AAG16878 Arabidopsis thalia
31	46	14.4	341	21	AAG21656 Arabidopsis thalia
32	46	14.4	351	23	ABB90981 Arabidopsis thalia
33	45	14.1	8	23	AAU12053 Aryl-sulfotransfer
34	45	14.1	273	23	ABB91902 Herbicidally activ
35	45	14.1	314	23	ABB93110 Herbicidally activ
36	45	14.1	318	21	AAG17287 Arabidopsis thalia
37	45	14.1	331	21	AAG17286 Arabidopsis thalia
38	45	14.1	331	23	ABB91710 Herbicidally activ
39	45	14.1	351	21	AAG17285 Arabidopsis thalia
40	44	13.8	8	19	AAW40501 Human EST protein
41	43	13.4	216	21	AAG04466 Arabidopsis thalia
42	43	13.4	338	21	AAG16972 Arabidopsis thalia
43	43	13.4	346	23	ABB91056 Herbicidally activ
44	43	13.4	350	21	AAG17409 Arabidopsis thalia
45	43	13.4	350	23	ABB91589 Herbicidally activ

ALIGNMENTS

RESULT 1
AAW23657
ID AAW23657 standard; Protein; 295 AA.
XX
AC AAW23657;
XX
DT 10-OCT-1997 (first entry)
XX
DE E6AP-binding protein cln25.
XX
KW Human; cell differentiation; survival; carcinoma; psoriasis;
KW ubiquitination; p53; tumour suppressor; homeostasis; papilloma virus;
KW epithelial cell; acne; ichthyosis; aphthous ulcer; hair growth;
KW antibody; cell proliferation.
XX
OS Homo sapiens.
XX
PN WO9640767-A2.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09040.
XX
PR 07-JUN-1995; 95US-0484878.
XX
PA (MITO-) MITOTIX INC.
XX
PI Beer-Romero PL, Draetta G, Rolfe M;
XX
DR WPI: 1997-087053/08.
DR N-PSDB; AAY78309.
XX
PT E6AP-binding proteins and related nucleic acid - useful for

modulating cell differentiation, survival etc., partic. for treatment and diagnosis of carcinoma, psoriasis, etc

Claim 1; Page 70-71; 83pp; English.

The present sequence represents the E6AP-binding protein c1n25. E6AP mediates ubiquitination and so the inactivation of e.g. p53 tumour suppressor. The E6AP-binding protein is likely to be involved in normal cell homeostasis and in the pathogenesis of proliferative and differentiation disorders, e.g. regulation of gene expression or the cell cycle, modification of cell surface receptors, biogenesis of ribosomes and DNA repair. The protein, which can optionally be generated in vivo by gene therapy, may be useful in treatment and prevention of papilloma virus infected transformed cells and carcinoma, and may also be used to regulate epithelial cell processes more generally, e.g. in cases of psoriasis, acne, ichthyosis, aphthous ulcers. It may also be used to control wound healing, and inhibit growth of hair. It can also be used to generate antibodies which are used in immunoassays to determine the protein levels. The encoding nucleic acid can be used to prepare recombinant proteins and oligonucleotides useful as probes and primers for diagnostic detection of (mutant) mRNA for the protein in (transformed) cells, also for antisense therapy. It can also be used for the detection of mutations in E6AP-binding protein-encoding genes, mis-expression of these genes or mis-incorporation of the protein in a transcription regulatory complex containing E6AP can be used to assess risk of disorders characterised by cell proliferation.

Sequence 295 AA;

Query Match 16.9%; Score 54; DB 18; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.075;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGGXGDKXXFT 232
||| |||||
Db 257 RKGAGDKTKTFT 269

RESULT 2

AAV67294
ID AAV67294 standard; Protein; 295 AA.

AC AAV67294;

DT 05-APR-2000 (first entry)

DE Human STP2 (phenol sulphotransferase 2) amino acid sequence.

KW Single nucleotide polymorphism; SNP; STP2; phenol sulphotransferase;
KW probe; genotyping; human; drug metabolism.

OS Homo sapiens.

PN WO9964630-A1.

PD 16-DEC-1999.

PF 09-JUN-1999; 99WO-US13094.

PR 10-JUN-1998; 98US-0088710.

PA (AXYS-) AXYS PHARM INC.

PI Guida M, Kurth J;

DR WPI; 2000-105892/09.

DR N-PSDB; AAZ59353.

XX Novel nucleic acid used for genotyping, e.g. to predict rate of drug
PT metabolism -

PS Disclosure; Page 28; 46pp; English.

CC This is the human phenol sulphotransferase 2 (STP2) protein sequence. The
CC invention relates to sequences AAZ59305-259352 which are fragments of
CC the STP2 gene. The fragments are from the 8 exons, the promoter region,
CC 3' and 5' untranslated regions of the STP2 gene. Each of the sequences
CC contains a newly identified STP2 gene single nucleotide polymorphism
CC (SNP). STP2 is a phenol sulphotransferase. Substrates for STP2 include
CC minoxidil, acetaminophen, and paracetamol. Several of the nucleotide
CC changes identified at the polymorphism sites, give rise to an amino acid
CC change. Amino acid changes may result in altered enzyme activity. The
CC sequences can be used as probes for detecting STP2 polymorphisms. The
CC polymorphic probes are used in screening and genotyping, i.e. to predict
CC the rate of metabolism of STP2 substrates, potential drug-drug
CC interactions and adverse side effects. They can also be used to detect
CC diseases resulting from accidental or occupational exposure to toxins
CC and to establish animal, cell or in vitro models for drug metabolism.

Sequence 295 AA;

Query Match 16.9%; Score 54; DB 21; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.075;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGGXGDKXXFT 232
||| |||||
Db 257 RKGAGDKTKTFT 269

RESULT 3

ABG27554

ID ABG27554 standard; Protein; 1305 AA.

AC ABG27554;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #27545.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS91741.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 57913; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (ii) and its binding partners are useful in medical
 CC imaging of sites expressing (ii). (i) and (ii) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 1305 AA;
 Query Match 16.9%; Score 54; DB 22; Length 1305;
 Best Local Similarity 69.2%; Pred. No. 0.35;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDKXXFT 232
 ||| |||||
 Db 554 RKGWAGDKTKTFT 566

RESULT 4
 AAB45369
 ID AAB45369 standard; Protein; 129 AA.
 XX
 AC AAB45369;
 XX
 DT 14-FEB-2001 (first entry)
 XX
 DE Gene 35 human secreted protein homologous amino acid sequence #121.
 XX
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; nontropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; vulnary; gene therapy; autoimmune disease;
 KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;
 KW cerebrovascular disorder; nervous system disorder; infection; skin aging;
 KW ocular disorder; wound healing; food additive; preservative.
 XX
 OS Homo sapiens.
 XX
 XX WO2000061628-A1.
 PN
 PD 19-OCT-2000.
 XX
 XX 06-APR-2000; 2000WO-US09070.
 XX
 XX 09-APR-1999; 99US-0128695.
 XX
 XX 14-JAN-2000; 2000US-0176052.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Ruben SM, Komatsoulis G;
 XX
 XX WPI; 2000-619228/59.
 XX
 XX New nucleic acid molecules encoding 49 human secreted proteins for
 XX diagnosing, preventing, treating or ameliorating medical conditions and
 XX used as food additives or preservatives -
 XX
 XX Disclosure; Page 443-444; 454pp; English.
 XX
 XX The polynucleotide sequences given in AAC81086 to AAC81134 encode the
 XX human secreted proteins given in AAB45308 to AAB45356. AAB45357 to
 XX AAB45384 represent human secreted polypeptide sequences and proteins
 XX homologous to them, which are given in the exemplification of the present
 XX invention. Human secreted proteins have activities based on the tissues
 XX and cells the genes are expressed in. Examples of activities include:
 XX antiarthritic; immunosuppressive; antirheumatic; antiproliferative;

CC cytostatic; cardiant; vasotropic; cerebroprotective; nontropic;
 CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
 CC and vulnary. The polynucleotides and polypeptides can be used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases, hyperproliferative disorders, cardiovascular
 CC disorders, cerebrovascular disorders, angiogenesis, nervous system
 CC disorders, infections caused by bacteria, viruses and fungi and ocular
 CC disorders. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional
 CC components. AAC81077 to AAC81085 and AAB45307 represent sequences used in
 CC the exemplification of the present invention.

XX
 SQ Sequence 129 AA;
 Query Match 16.2%; Score 52; DB 21; Length 129;
 Best Local Similarity 69.2%; Pred. No. 0.075;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDKXXFT 232
 ||| |||||
 Db 91 RKGITGDWKNHFT 103

RESULT 5
 AAB45370
 ID AAB45370 standard; Protein; 129 AA.
 XX
 AC AAB45370;
 XX
 DT 14-FEB-2001 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 35 SEQ ID NO:122.
 XX
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; nontropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; vulnary; gene therapy; autoimmune disease;
 KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;
 KW cerebrovascular disorder; nervous system disorder; infection; skin aging;
 KW ocular disorder; wound healing; food additive; preservative.
 XX
 OS Homo sapiens.
 XX
 XX WO2000061628-A1.
 PN
 PD 19-OCT-2000.
 XX
 XX 06-APR-2000; 2000WO-US09070.
 XX
 XX 09-APR-1999; 99US-0128695.
 XX
 XX 14-JAN-2000; 2000US-0176052.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Ruben SM, Komatsoulis G;
 XX
 XX WPI; 2000-619228/59.
 XX
 XX New nucleic acid molecules encoding 49 human secreted proteins for
 XX diagnosing, preventing, treating or ameliorating medical conditions and
 XX used as food additives or preservatives -
 XX
 XX Disclosure; Page 444; 454pp; English.
 XX
 XX The polynucleotide sequences given in AAC81086 to AAC81134 encode the

CC human secreted proteins given in AAB45308 to AAB45356. AAB45357 to
 CC AAB45384 represent human secreted polypeptide sequences and proteins
 CC homologous to them, which are given in the exemplification of the present
 CC invention. Human secreted proteins have activities based on the tissues
 CC and cells the genes are expressed in. Examples of activities include:
 CC antitumor; immunosuppressive; antirheumatic; antiproliferative;
 CC cytoprotective; cardiant; vasotropic; cerebroprotective; neurotropic;
 CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
 CC and vulnary. The polynucleotides and polypeptides can be used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases, hyperproliferative disorders, cardiovascular
 CC disorders, cerebrovascular disorders, angiogenesis, nervous system
 CC disorders, infections caused by bacteria, viruses and fungi and ocular
 CC disorders. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional
 CC components. AAC81077 to AAC81085 and AAB45307 represent sequences used in
 CC the exemplification of the present invention.

SQ Sequence 129 AA;

Query Match 16.2%; Score 52; DB 21; Length 129;
 Best Local Similarity 69.2%; Pred. No. 0.075;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMWXXFT 232
 DB 91 RKGITGDWKNHFT 103

RESULT 6
 AAG75324

ID AAG75324 standard; Protein; 141 AA.

XX AAG75324;

DT 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6088.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 XX colorectal carcinoma; chromosome 2.

XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX N-PSDB; AAB34729.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 XX useful for preventing, diagnosing and/or treating colorectal cancers -

PS Claim 11; Page 7554; 9803pp; English.

XX

CC AAB32943 to AAB37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P.
 CC Inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAB37196 to AAB37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

SQ Sequence 141 AA;

Query Match 16.2%; Score 52; DB 22; Length 141;
 Best Local Similarity 69.2%; Pred. No. 0.083;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMWXXFT 232
 DB 103 RKGTVGDWKNHFT 115

RESULT 7

AAB40498

ID AAB40498 standard; Protein; 294 AA.

XX AAB40498;

DT 14-JUL-1998 (first entry)

XX Human EST protein.

XX Oestrogen sulphotransferase; EST; drug metabolism; xenobiotic; hormone;
 XX neurotransmitter; sulphate conjugation; renal excretion; screening;
 XX antigen; immunohistochemistry.

XX Homo sapiens.

XX US5744355-A.

XX 28-APR-1998.

XX 09-MAY-1995; 95US-0437795.

XX 09-MAY-1995; 95US-0437795.

XX 18-OCT-1994; 94US-0325562.

XX (MAYO-) MAYO FOUNDATION.

XX Aksoy IA, Weinshilboum RM, Wood TC;

XX WPI; 1998-271072/24.

XX N-PSDB; AAV11154.

XX Cells and vectors encoding human oestrogen sulphoto-transferase - are
 XX useful for the recombinant expression of the enzyme

PS Claim 2; Fig 2; 30pp; English.

XX This sequence represents a novel human oestrogen sulphotransferase (EST).
 CC This protein can be inserted into cell lines or vectors which are useful
 CC for the recombinant expression of the enzyme. Sulphotransferases are
 CC involved in the metabolism of drugs, xenobiotics, neurotransmitters and
 CC hormones by sulphate conjugation. This is required for increase in water

CC solubility (leading to renal excretion) and in some cases for activation
 CC of these compounds. The recombinant products are useful for the
 CC determination of hormones and drugs which are metabolised by the enzyme
 CC and screening of compounds structurally similar to oestrogen that can be
 CC metabolised by the in vivo sulphotransferase and their affinity for the
 CC enzyme. DNA encoding the enzyme can be used to quantify expression levels
 CC in different tissues and detect mutations. The enzyme can also be used
 CC as an antigen to raise antibodies for immunohistochemical analysis.
 XX
 SQ Sequence 294 AA;

Query Match 16.2%; Score 52; DB 19; Length 294;
 Best Local Similarity 69.2%; Pred. No. 0.18; Mismatches 4; Indels 0; Gaps 0;
 Matches 9; Conservative 0;

Qy 220 RKGXXGDMKXFT 232
 Db 256 RKGITGDMKXFT 268

RESULT 8
 AAW4247
 ID AAW4247 standard; Protein; 294 AA.
 XX
 AC AAW4247;

XX 15-MAY-1998 (first entry)
 DT Human oestrogen sulphotransferase.
 DE Human; oestrogen sulphotransferase; estrogen sulfotransferase;
 KW Liver EST; biotransformation.
 KW Homo sapiens.
 OS
 XX US5714594-A.
 PN 03-FEB-1998.
 PD 18-OCT-1994; 94US-0325562.
 PF 18-OCT-1994; 94US-0325562.
 PR (MAYO-) MAYO FOUNDATION.
 XX Aksoy IA, Weinshilboun RM, Wood TC;
 XX WPI; 1998-144284/13.
 XX N-PSDB; AAV12466.

XX DNA encoding human oestrogen sulphotransferase - useful for research
 PT into oestrogen biotransformation
 XX Claim 1; Column 21-24; 25pp; English.
 XX The present sequence represents a human oestrogen sulphotransferase
 CC (EST) protein. The isolation and expression of cDNA which encodes human
 CC liver EST enables the biotransformation of oestrogens to be studied.
 XX
 SQ Sequence 294 AA;
 Query Match 16.2%; Score 52; DB 19; Length 294;
 Best Local Similarity 69.2%; Pred. No. 0.18; Mismatches 4; Indels 0; Gaps 0;
 Matches 9; Conservative 0;

Qy 220 RKGXXGDMKXFT 232
 Db 256 RKGITGDMKXFT 268

RESULT 9
 AAW40502
 ID AAW40502 standard; Protein; 13 AA.

XX AAW40502;
 AC 14-JUL-1998 (first entry)
 DT Human EST protein conserved peptide motif #4.
 DE Oestrogen sulphotransferase; EST; drug metabolism; xenobiotic; hormone;
 XX neurotransmitter; sulphate conjugation; renal excretion; screening;
 KW antigen; immunohistochemistry.
 KW Homo sapiens.
 OS

Key Location/Qualifiers
 FT Misc-difference 4 /label= any amino acid
 FT Misc-difference 5 /label= any amino acid
 FT Misc-difference 11 /label= any amino acid
 FT

US5744355-A.
 28-APR-1998.
 09-MAY-1995; 95US-0437795.
 09-MAY-1995; 95US-0437795.
 18-OCT-1994; 94US-0325562.

(MAYO-) MAYO FOUNDATION.

Aksoy IA, Weinshilboun RM, Wood TC;
 WPI; 1998-271072/24.

Cells and vectors encoding human oestrogen sulpho-transferase - are
 useful for the recombinant expression of the enzyme
 Example 1; Column 20; 30pp; English.

This sequence represents a motif found in a novel human oestrogen
 sulphotransferase (EST). This protein can be inserted into cell lines or
 vectors which are useful for the recombinant expression of the enzyme.
 CC Sulphotransferases are involved in the metabolism of drugs, xenobiotics,
 CC neurotransmitters and hormones by sulphate conjugation. This is required
 CC for increase in water solubility (leading to renal excretion) and in some
 CC cases for activation of these compounds. The recombinant products are
 CC useful for the determination of hormones and drugs which are metabolised
 CC by the enzyme and screening of compounds structurally similar to
 CC oestrogen that can be metabolised by the in vivo sulphotransferase and
 CC their affinity for the enzyme. DNA encoding the enzyme can be used to
 CC quantify expression levels in different tissues and detect mutations. The
 CC enzyme can also be used as an antigen to raise antibodies for
 CC immunohistochemical analysis.

Sequence 13 AA;
 Query Match 15.9%; Score 51; DB 19; Length 13;
 Best Local Similarity 92.3%; Pred. No. 0.011;
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 220 RKGXXGDMKXFT 232
 Db 1 RKGXXGDMKXFT 13

RESULT 10
 AAU78272
 ID AAU78272 standard; Peptide; 13 AA.
 XX
 AC AAU78272;
 XX

05-JUN-2002 (first entry)

SULTn PAPS 3' phosphate binding site consensus sequence #2.

Nervous system-specific sulfotransferase; SULTn; sulfonation; neurological disorder; multiple sclerosis; Huntington's chorea; nervous system-mediated muscular atrophy; amyotrophic lateral sclerosis; myasthenia gravis; Alzheimer's disease; senile dementia; Schizophrenia; bipolar disorder; cytosolic sulfotransferase gene family; cerebellar ataxia; PAPS 3' phosphate binding site.

Synthetic.

Key Location/Qualifiers

Misc-difference 4..5 /label= unknown

Misc-difference 10..11 /label= unknown

WO200218541-A2.

07-MAR-2002.

04-SEP-2001; 2001WO-US41986.

01-SEP-2000; 2000US-229929P.

(UYBO-) UNIV BOSTON.

Farb DH, Martin S;

WPI; 2002-281063/32.

New nervous system-specific sulfotransferase (SULTn) polypeptide for treating neurological disorders including multiple sclerosis, nervous system-mediated muscular atrophy and Huntington's chorea -

Disclosure; Fig 1d; 46pp; English.

The present invention relates to a new nervous system-specific sulfotransferase (SULTn) polypeptide. The SULTn polypeptide of the invention comprises a nervous system-specific sulfotransferase, comprising a sequence of 284 amino acids, given in the specification. The invention is useful for identifying an effector of nervous system-specific sulfonation by contacting the polypeptide with a candidate effector compound and assaying the ability of the candidate effector compound to effect the sulfotransferase function of the polypeptide. The invention is also useful for inhibiting nervous system-specific sulfonation by contacting the polypeptide under physiological conditions with a compound capable of interacting with and inhibiting the sulfonation activity of the polypeptide, where the compound is an antibody, peptide, polypeptide, nucleic acid, organic molecule or an inorganic molecule. The polypeptide of the invention is also useful for treating a neurological disorder, including multiple sclerosis, nervous system-mediated muscular atrophy, amyotrophic lateral sclerosis, Huntington's chorea, myasthenia gravis, Alzheimer's disease, senile dementia, Schizophrenia, bipolar disorder and cerebellar ataxia. The present amino acid sequence represents the SULTn protein PAPS 3', cytosolic sulfotransferase gene family.

Sequence 13 AA;

Query Match 15.9%; Score 51; DB 23; Length 13;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 220 RKGXGDKXXFT 232

Db 1 RKGXGDKXXFT 13

RESULT 11

AAU12052

ID AAU12052 standard; protein; 23 AA.

XX

AC AAU12052;

XX

DT 09-APR-2002 (first entry)

XX

DE Sulfotransferase Region I-Region IV consensus sequence.

XX

KW Plant; transgenic; marine eelgrass; zosteric acid biosynthesis;

KW saline-resistance; anoxia-resistance; anti-fouling genetic trait;

KW marine vascular plant; sulphated phenolic compound; Zostera marina;

KW sulfotransferase; ST; enzyme.

XX

OS Synthetic.

XX

XX

XX Key Location/Qualifiers

XX 1..9

FT Region

FT /label= Region_I

FT /note= "Region I is separated from Region IV by 190-210 amino acid residues"

FT

FT Misc-difference 8

FT /label= Asn, Thr

FT

FT Misc-difference 10

FT

FT /label= Xaa190-210

FT /note= "Xaa represents any amino acid and Xaa190-210 represents the 190 to 210 amino acid residues intervening between conserved Region I and conserved Region IV residues"

FT

FT Region 11..23

FT /label= Region_IV

FT /note= "Region IV is separated from Region I by 190-210 amino acid residues"

FT

FT Misc-difference 14..15

FT /note= "Any amino acid"

FT

FT Misc-difference 20..21

FT /note= "Any amino acid"

XX

XX

PN WO200185971-A2.

XX

XX

PD 15-NOV-2001.

XX

XX

PF 10-MAY-2001; 2001WO-US15412.

XX

XX

PR 10-MAY-2000; 2000US-202529P.

XX

XX (PHYC-) PHYCOGEN INC.

PA

XX

XX

PI Alberte RS, Smith RD;

XX

XX

DR WPI; 2002-121947/16.

XX

XX

XX New transgenic plants comprising a zosteric acid biosynthetic gene, a saline resistance gene or a hypoxia resistance gene derived from Zostera marina, useful for producing plants with antifouling traits -

XX

PS Disclosure; Page 29; 117pp; English.

XX

XX

CC The present invention relates to a new transgenic plant comprising a heterologous gene derived from the marine eelgrass Zostera marina or at least one heterologous nucleotide sequence encoding a zosteric acid biosynthetic function, a saline-resistance function, or a anoxia-resistance function. The invention describes the method of producing a transgenic plant possessing an anti-fouling genetic trait by providing a cDNA population derived from a marine vascular plant, isolating from the cDNA population a nucleic acid species which hybridises to a nucleic acid that encodes a sulfotransferase (ST), an alcohol dehydrogenase (ADH), phenylalanine ammonia lyase (PAL) or a cinnamate-4-hydroxylase (CH), and transforming a target host plant with the isolated nucleic acid. The plant is useful in the genetic engineering of plant species having desirable genetic traits such as antifouling traits, salt and anoxia resistance, and pathogen defence strategy. The expression of such biosynthetic enzymes are sufficient to support the

CC production of zosteric acid and other sulphated phenolic compounds in
CC a target plant. The present sequence represents Region I-Region IV
CC consensus sequence found in almost all sulfotransferase proteins.

XX SQ Sequence 23 AA;
Query Match 15.9%; Score 51; DB 23; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 220 RKGXXGDMXXFT 232
DB 11 RKGXXGDMXXFT 23
|||||

RESULT 12
AAG03995
ID AAG03995 standard; Protein; 104 AA.
XX
AC AAG03995;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 8076.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR N-PSDB; AAC04001.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 8076; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
XX were prepared from total human RNAs or polyA+ RNAs derived from 30
XX different tissues. EST sequences usually correspond mainly to the 3'
XX untranslated region (UTR) of the mRNA because they are often obtained
XX from oligo-dr primed cDNA libraries. Such ESTs are not well suited for
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in
XX those cases where longer cDNA sequences have been obtained, the full 5'
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
XX ends and can therefore be used to obtain full length cDNAs and genomic
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
XX chromosome mapping procedures. They are used to obtain upstream
XX regulatory sequences and to design expression and secretion vectors.

XX SQ Sequence 104 AA;
Query Match 15.6%; Score 50; DB 21; Length 104;
Best Local Similarity 88.9%; Pred. No. 0.14;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TYPKSGTGW 9
|||||

Db 44 TYPKSGTTW 52
RESULT 13
ABB81794
ID ABB81794 standard; Protein; 283 AA.
XX
AC ABB81794;
XX
DT 24-SEP-2002 (first entry)
XX
DE Human sulfotransferase #2.
XX
KW Human; sulfotransferase; cytostatic; gene therapy; cancer; COPD;
KW chronic obstructive pulmonary disease.

XX OS Homo sapiens.
XX
PN WO200253713-A2.
XX
PD 11-JUL-2002.
XX
PF 21-DEC-2001; 2001WO-EPI5233.
XX
PR 02-JAN-2001; 2001US-258812P.
XX
PA (FARB) BAYER AG.
XX
PI Koehler RH;
XX
XX WPI; 2002-528859/56.
DR N-PSDB; ABN83171.
XX

XX Novel human sulfotransferase polypeptide useful in therapeutic methods
PT for treating disorders such as cancer or chronic obstructive pulmonary
PT disease, and for screening for human sulfotransferase.modulators -
XX
PS Disclosure; Fig 6; 103pp; English.
XX
CC The invention relates to a novel human sulfotransferase polypeptide
CC The sulfotransferase has cytostatic activity. The polynucleotide of the
CC invention may have a use in gene therapy. The polypeptide and of the
CC polynucleotide of the invention are useful in preventing, ameliorating,
CC or correcting diseases including cancer and chronic obstructive pulmonary
CC disease (COPD). The sequence represents a human sulfotransferase of the
CC invention.
XX
SQ Sequence 283 AA;

Query Match 15.6%; Score 50; DB 23; Length 283;
Best Local Similarity 88.9%; Pred. No. 0.4;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TYPKSGTGW 9
DB 32 TYPKSGTTW 40
|||||

RESULT 14
AAE12208
ID AAE12208 standard; Protein; 304 AA.
XX
AC AAE12208;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human ST drug-metabolising protein 1 encoded by DNA transcript 1.

XX Human; sulfotransferase; drug-metabolising enzyme mediated disorder;
KW ST; DME; toxicity study; gene therapy.
XX
OS Homo sapiens.
XX

FH Key Location/Qualifiers
 FT Modified-site 173..175
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT Modified-site 180..183
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT Modified-site 236..239
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT Modified-site 248..251
 FT Modified-site /note= "Asn is N-glycosylated"
 FT Modified-site 258..261
 FT Modified-site /note= "Asn is N-glycosylated"
 FT Modified-site 295..300
 FT Modified-site /note= "N-myristoylation site"
 FT
 XX WO200172977-A2
 PN
 XX
 XX 04-OCT-2001.
 XX
 XX 27-MAR-2001; 2001WO-US09659.
 XX
 XX 27-MAR-2000; 2000US-192408P.
 PR 20-JUN-2000; 2000US-212725P.
 PR 03-JUL-2000; 2000US-0609816.
 XX
 XX (PEKE) PE CORP NY.
 XX
 XX Woodage T, Wei M, Kodira C, Di Francesco V, Beasley EM;
 PI WPI; 2001-616484/71.
 DR N-PSDB; AAD19776.
 DR
 XX
 XX New polypeptides useful in labeling reagents with tagged sulfur and
 PT identifying modulators of the protein comprises proteins related to
 FT sulfotransferase drug-metabolizing enzyme subfamily
 XX
 XX Claim 1; Fig 2; 83pp; English.
 PS
 XX The invention relates to sulfotransferase (ST) drug-metabolising
 CC enzyme (DME) peptides, proteins and nucleic acid molecules encoding
 CC them. Sequences of the invention are useful in labelling reagents
 CC with tagged sulphur, for the development of human therapeutics, as
 CC query sequences to perform a search against sequence databases, for
 CC identifying other family members or related sequences, to raise
 CC antibodies or to elicit another immune response, as a reagent in
 CC assays designed to quantitatively determine levels of the protein
 CC in biological fluids, to metabolise compounds for toxicity studies,
 CC and as markers for tissues in which the corresponding protein is
 CC preferentially expressed. Polypeptides of the invention is also
 CC useful in diagnosing a disease, predisposition to a disease mediated
 CC by the peptide, in pharmacogenomic analysis and treating a disorder
 CC characterised by absence of inappropriate or unwanted expression of
 CC the protein. Polynucleotides of the invention are useful for
 CC constructing recombinant vectors and host cells, expressing antigenic
 CC portions of the proteins, in genotyping and mapping experiments, for
 CC designing ribozymes, for monitoring the effectiveness of modulating
 CC compounds on the expression or activity of the protein, in diagnostic
 CC assays, antisense constructs to control protein expression and in
 CC gene therapy. The present sequence is ST drug-metabolising protein 1
 CC encoded by DNA transcript 1.
 XX
 SQ Sequence 304 AA;
 Query Match 15.6%; Score 50; DB 22; Length 304;
 Best Local Similarity 88.9%; Pred. No. 0.43;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 TYPKSGTXW 9
 |||||
 Db 53 TYPKSGTGW 61
 RESULT 15
 AAE12209

ID
 XX
 AC AAE12209 standard; Protein; 304 AA.
 AC AAE12209;
 DT 18-DEC-2001 (first entry)
 XX
 DE Human ST drug-metabolising protein 2 encoded by DNA transcript 2.
 XX
 XX Human; sulfotransferase; drug-metabolising enzyme mediated disorder;
 KW ST; DME; toxicity study; Gene therapy.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT Modified-site 173..175
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT Modified-site 180..183
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT Modified-site 236..239
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT Modified-site 248..251
 FT Modified-site /note= "Asn is N-glycosylated"
 FT Modified-site 255..258
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT Modified-site 295..300
 FT Modified-site /note= "N-myristoylation site"
 FT Modified-site 299..301
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT
 XX WO200172977-A2.
 PN
 XX
 XX 04-OCT-2001.
 XX
 XX 27-MAR-2001; 2001WO-US09659.
 XX
 XX 27-MAR-2000; 2000US-192408P.
 PR 20-JUN-2000; 2000US-212725P.
 PR 03-JUL-2000; 2000US-0609816.
 XX
 XX (PEKE) PE CORP NY.
 XX
 XX Woodage T, Wei M, Kodira C, Di Francesco V, Beasley EM;
 PI WPI; 2001-616484/71.
 DR N-PSDB; AAD19777.
 DR
 XX
 XX New polypeptides useful in labeling reagents with tagged sulfur and
 PT identifying modulators of the protein comprises proteins related to
 FT sulfotransferase drug-metabolizing enzyme subfamily
 XX
 XX Claim 1; Fig 2; 83pp; English.
 PS
 XX The invention relates to sulfotransferase (ST) drug-metabolising
 CC enzyme (DME) peptides, proteins and nucleic acid molecules encoding
 CC them. Sequences of the invention are useful in labelling reagents
 CC with tagged sulphur, for the development of human therapeutics, as
 CC query sequences to perform a search against sequence databases, for
 CC identifying other family members or related sequences, to raise
 CC antibodies or to elicit another immune response, as a reagent in
 CC assays designed to quantitatively determine levels of the protein
 CC in biological fluids, to metabolise compounds for toxicity studies,
 CC and as markers for tissues in which the corresponding protein is
 CC preferentially expressed. Polypeptides of the invention is also
 CC useful in diagnosing a disease, predisposition to a disease mediated
 CC by the peptide, in pharmacogenomic analysis and treating a disorder
 CC characterised by absence of inappropriate or unwanted expression of
 CC the protein. Polynucleotides of the invention are useful for
 CC constructing recombinant vectors and host cells, expressing antigenic
 CC portions of the proteins, in genotyping and mapping experiments, for
 CC designing ribozymes, for monitoring the effectiveness of modulating
 CC compounds on the expression or activity of the protein, in diagnostic
 CC assays, antisense constructs to control protein expression and in
 CC gene therapy. The present sequence is ST drug-metabolising protein 2
 CC encoded by DNA transcript 2.

```
CC encoded by DNA transcript 2.
XX
SQ Sequence 304 AA;
Query Match 15.6%; Score 50; DB 22; Length 304;
Best Local Similarity 88.9%; Pred. No. 0.43;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TYPKSGTYW 9
   |||||
Db 53 TYPKSGTWW 61

Search completed: June 5, 2003, 12:29:25
Job time : 72 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:28:13 ; Search time 27 Seconds
(without alignments)
252.819 Million cell updates/sec

Title: US-09-854-122-1
Perfect score: 320
Sequence: 1 TYPKSGTXXXXXXX.....XXXXXXXXXGDKXXFT 232

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/iaa/5A_COMB pep.*
2: /cgn2_6/prodata/1/iaa/5B_COMB pep.*
3: /cgn2_6/prodata/1/iaa/6A_COMB pep.*
4: /cgn2_6/prodata/1/iaa/6B_COMB pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1 pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	54	16.9	295 4	US-09-328-174A-3
2	52	16.2	294 1	US-08-325-562-2
3	52	16.2	294 1	US-08-437-795-2
4	51	15.9	13 1	US-08-325-562-17
5	51	15.9	13 1	US-08-437-795-17
6	50	15.6	283 4	US-09-609-816-8
7	50	15.6	283 4	US-09-609-816-9
8	50	15.6	283 4	US-09-609-816-10
9	50	15.6	295 3	US-09-150-133-13
10	50	15.6	295 3	US-09-150-141-13
11	50	15.6	295 4	US-09-374-493-13
12	50	15.6	295 4	US-09-374-824-13
13	50	15.6	295 4	US-09-374-492-13
14	50	15.6	304 4	US-09-609-816-4
15	50	15.6	304 4	US-09-609-816-5
16	50	15.6	304 4	US-09-609-816-6
17	44	13.8	8 1	US-08-325-562-16
18	44	13.8	8 1	US-08-437-795-16
19	44	13.8	36 4	US-09-167-681-36
20	44	13.8	37 4	US-09-167-681-44
21	44	13.8	37 4	US-09-167-681-52
22	38	11.9	7 3	US-09-150-133-44
23	38	11.9	7 3	US-09-150-141-44
24	38	11.9	7 4	US-09-374-493-44
25	38	11.9	7 4	US-09-374-824-44
26	38	11.9	7 4	US-09-374-492-44
27	37	11.6	32 1	US-08-477-727A-72

28	37	11.6	32	1	US-08-477-727A-74	Sequence 74, Appl
29	36	11.2	350	2	US-08-828-922-1	Sequence 1, Appl
30	35	10.9	42	1	US-08-062-472B-44	Sequence 44, Appl
31	35	10.9	42	2	US-08-835-231-11	Sequence 11, Appl
32	35	10.9	42	4	US-09-108-661-11	Sequence 11, Appl
33	35	10.9	42	4	US-09-260-846-21	Sequence 21, Appl
34	35	10.9	187	2	US-08-835-231-16	Sequence 16, Appl
35	35	10.9	187	4	US-09-108-661-16	Sequence 16, Appl
36	35	10.9	350	2	US-08-852-481-2	Sequence 2, Appl
37	34	10.6	50	4	US-09-167-681-38	Sequence 38, Appl
38	33	10.3	11	1	US-08-445-745-6	Sequence 6, Appl
39	33	10.3	11	1	US-08-445-745-10	Sequence 10, Appl
40	33	10.3	11	4	US-08-456-466-3	Sequence 3, Appl
41	33	10.3	11	4	US-08-456-466-7	Sequence 7, Appl
42	33	10.3	117	2	US-08-888-497-44	Sequence 44, Appl
43	33	10.3	117	4	US-09-362-230-44	Sequence 44, Appl
44	33	10.3	117	5	PCT-US94-07926-44	Sequence 44, Appl
45	33	10.3	118	2	US-08-888-497-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-09-328-174A-3
; Sequence 3, Application US/09328174A
; Patent No. 6448003
; GENERAL INFORMATION:
; APPLICANT: Guida, Marco
; APPLICANT: Kurth, Janice
; TITLE OF INVENTION: Genotyping Human Phenol Sulfotransferase
; FILE REFERENCE: 4389-6 (formerly SEQ-16P)
; CURRENT APPLICATION NUMBER: US/09/328,174A
; CURRENT FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 09/328,174
; PRIOR FILING DATE: 1999-06-08
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-328-174A-3

Query Match 16.9%; Score 54; DB 4; Length 295;
Best Local Similarity 69.2%; Pred No. 0.058;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDKXXFT 232
||| |||||
Db 257 RKGAGDKWKTFT 269

RESULT 2
US-08-325-562-2
; Sequence 2, Application US/08325562
; Patent No. 5714594
; GENERAL INFORMATION:
; APPLICANT: Weinshilboum, Richard M.
; APPLICANT: Aksoy, Ibrahim A.
; APPLICANT: Wood, Thomas C.
; TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN
; TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:

10

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Schwegman, Lundberg & Woessner P.A.
 STREET: 3500 IDS Center
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/437,795
 FILING DATE: 09-MAY-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Muetting, Ann M.
 REGISTRATION NUMBER: 33,977
 REFERENCE/DOCKET NUMBER: 150.137US1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-339-0331
 TELEFAX: 612-339-3061
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 13 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-437-795-17

Query Match 15.9%; Score 51; DB 1; Length 13;
 Best Local Similarity 92.3%; Pred. No. 0.0087;
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 220 RKGXXGDKXXFT 232
 Db 1 RKGXXGDKXXFT 13

RESULT 6
 US-09-609-816-8
 ; Sequence 8, Application US/09609816
 ; Patent No. 6436684
 ; GENERAL INFORMATION:
 ; APPLICANT: Woodage, Trevor
 ; APPLICANT: Wei, Minh Hui
 ; APPLICANT: Kodira, Chinappa
 ; APPLICANT: Beasley, Ellen
 ; APPLICANT: DiFrancesco, Valentina
 ; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
 ; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
 ; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
 ; FILE REFERENCE: CL000669PCT
 ; CURRENT APPLICATION NUMBER: US/09/609,816
 ; CURRENT FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/192,408
 ; PRIOR FILING DATE: 2000-03-27
 ; PRIOR APPLICATION NUMBER: 60/212,725
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: 09/609,816
 ; PRIOR FILING DATE: 2000-07-03
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 283
 ; TYPE: PRT
 ; ORGANISM: HUMAN
 US-09-609-816-8

Query Match 15.6%; Score 50; DB 4; Length 283;
 Best Local Similarity 69.2%; Pred. No. 0.27;

Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 220 RKGXXGDKXXFT 232
 Db 245 RKGMPGDWKNYFT 257

RESULT 7
 US-09-609-816-9
 ; Sequence 9, Application US/09609816
 ; Patent No. 6436684
 ; GENERAL INFORMATION:
 ; APPLICANT: Woodage, Trevor
 ; APPLICANT: Wei, Minh Hui
 ; APPLICANT: Kodira, Chinappa
 ; APPLICANT: Beasley, Ellen
 ; APPLICANT: DiFrancesco, Valentina
 ; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
 ; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
 ; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
 ; FILE REFERENCE: CL000669PCT
 ; CURRENT APPLICATION NUMBER: US/09/609,816
 ; CURRENT FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/192,408
 ; PRIOR FILING DATE: 2000-03-27
 ; PRIOR APPLICATION NUMBER: 60/212,725
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: 09/609,816
 ; PRIOR FILING DATE: 2000-07-03
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9
 ; LENGTH: 283
 ; TYPE: PRT
 ; ORGANISM: HUMAN
 US-09-609-816-9

Query Match 15.6%; Score 50; DB 4; Length 283;
 Best Local Similarity 69.2%; Pred. No. 0.27;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXXGDKXXFT 232
 Db 245 RKGMPGDWKNYFT 257

RESULT 8
 US-09-609-816-10
 ; Sequence 10, Application US/09609816
 ; Patent No. 6436684
 ; GENERAL INFORMATION:
 ; APPLICANT: Woodage, Trevor
 ; APPLICANT: Wei, Minh Hui
 ; APPLICANT: Kodira, Chinappa
 ; APPLICANT: Beasley, Ellen
 ; APPLICANT: DiFrancesco, Valentina
 ; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
 ; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
 ; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
 ; FILE REFERENCE: CL000669PCT
 ; CURRENT APPLICATION NUMBER: US/09/609,816
 ; CURRENT FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/192,408
 ; PRIOR FILING DATE: 2000-03-27
 ; PRIOR APPLICATION NUMBER: 60/212,725
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: 09/609,816
 ; PRIOR FILING DATE: 2000-07-03
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10
 ; LENGTH: 283
 ; TYPE: PRT

US-09-374-493-13

Sequence 13, Application US/09374493

Patent No. 6204016

GENERAL INFORMATION:

APPLICANT: The Board of Regents of the University of Oklahoma

TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF

FILE REFERENCE: 5820.546

CURRENT APPLICATION NUMBER: US/09/374,493

CURRENT FILING DATE: 1999-08-13

EARLIER APPLICATION NUMBER: 09/150,133

EARLIER FILING DATE: 1998-09-09

EARLIER APPLICATION NUMBER: 60/072,994

EARLIER FILING DATE: 1998-01-29

EARLIER APPLICATION NUMBER: PCT/US99/16750

EARLIER FILING DATE: 1999-07-23

NUMBER OF SEQ ID NOS: 45

SOFTWARE: Wordperfect 8.0 (saved in ASCII format)

SEQ ID NO 13

LENGTH: 295

TYPE: PRT

ORGANISM: Mus musculus

FEATURE:

OTHER INFORMATION: estrogen sulfotransferase

US-09-374-493-13

Query Match 15.6%; Score 50; DB 4; Length 295;

Best Local Similarity 88.9%; Pred. No. 0.28;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TYPKSGTGW 9

DB 45 TYPKSGTGW 53

RESULT 12

US-09-374-824-13

Sequence 13, Application US/09374824

Patent No. 6207414

GENERAL INFORMATION:

APPLICANT: The Board of Regents of the University of Oklahoma

TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF

FILE REFERENCE: 5820.547

CURRENT APPLICATION NUMBER: US/09/374,824

CURRENT FILING DATE: 1999-08-13

EARLIER APPLICATION NUMBER: 09/150,133

EARLIER FILING DATE: 1998-09-09

EARLIER APPLICATION NUMBER: 60/072,994

EARLIER FILING DATE: 1998-01-29

EARLIER APPLICATION NUMBER: PCT/US99/16750

EARLIER FILING DATE: 1999-07-23

NUMBER OF SEQ ID NOS: 45

SOFTWARE: Wordperfect 8.0 (saved in ASCII format)

SEQ ID NO 13

LENGTH: 295

TYPE: PRT

ORGANISM: Mus musculus

FEATURE:

OTHER INFORMATION: estrogen sulfotransferase

US-09-374-824-13

Query Match 15.6%; Score 50; DB 4; Length 295;

Best Local Similarity 88.9%; Pred. No. 0.28;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TYPKSGTGW 9

DB 45 TYPKSGTGW 53

RESULT 13

US-09-374-492-13

Sequence 13, Application US/09374492

Patent No. 6207432

GENERAL INFORMATION:

APPLICANT: The Board of Regents of the University of Oklahoma

TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF

FILE REFERENCE: 5820.546

CURRENT APPLICATION NUMBER: US/09/374,492

CURRENT FILING DATE: 1999-08-13

EARLIER APPLICATION NUMBER: 09/150,133

EARLIER FILING DATE: 1998-09-09

EARLIER APPLICATION NUMBER: 60/072,994

EARLIER FILING DATE: 1998-01-29

EARLIER APPLICATION NUMBER: PCT/US99/16750

EARLIER FILING DATE: 1999-07-23

NUMBER OF SEQ ID NOS: 45

SOFTWARE: Wordperfect 8.0 (saved in ASCII format)

SEQ ID NO 13

LENGTH: 295

TYPE: PRT

ORGANISM: Mus musculus

FEATURE:

OTHER INFORMATION: estrogen sulfotransferase

US-09-374-492-13

Query Match 15.6%; Score 50; DB 4; Length 295;

Best Local Similarity 88.9%; Pred. No. 0.28;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TYPKSGTGW 9

DB 45 TYPKSGTGW 53

US-09-609-816-10

Sequence 13, Application US/09150133B

Patent No. 6060295

GENERAL INFORMATION:

APPLICANT: The Board of Regents of the University of Oklahoma

TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING

FILE REFERENCE: 5820.504

CURRENT APPLICATION NUMBER: US/09/150,133B

CURRENT FILING DATE: 1998-09-09

NUMBER OF SEQ ID NOS: 45

SOFTWARE: Wordperfect 5.1 (saved in ASCII format)

SEQ ID NO 13

LENGTH: 295

TYPE: PRT

ORGANISM: Mus musculus

FEATURE:

OTHER INFORMATION: estrogen sulfotransferase

US-09-150-133-13

Query Match 15.6%; Score 50; DB 3; Length 295;

Best Local Similarity 88.9%; Pred. No. 0.28;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TYPKSGTGW 9

DB 45 TYPKSGTGW 53

RESULT 10

US-09-150-141-13

Sequence 13, Application US/09150141B

Patent No. 6071732

GENERAL INFORMATION:

APPLICANT: The Board of Regents of the University of Oklahoma

TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING

FILE REFERENCE: 5820.495

CURRENT APPLICATION NUMBER: US/09/150,141B

CURRENT FILING DATE: 1998-09-09

NUMBER OF SEQ ID NOS: 45

SOFTWARE: Wordperfect 5.1 (saved in ASCII format)

SEQ ID NO 13

LENGTH: 295

TYPE: PRT

ORGANISM: Mus musculus

FEATURE:

OTHER INFORMATION: estrogen sulfotransferase

US-09-150-141-13

Query Match 15.6%; Score 50; DB 3; Length 295;

Best Local Similarity 88.9%; Pred. No. 0.28;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TYPKSGTGW 9

DB 45 TYPKSGTGW 53

RESULT 11

US-09-150-141-13

Sequence 13, Application US/09150141B

Patent No. 6071732

GENERAL INFORMATION:

APPLICANT: The Board of Regents of the University of Oklahoma

TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING

FILE REFERENCE: 5820.495

CURRENT APPLICATION NUMBER: US/09/150,141B

CURRENT FILING DATE: 1998-09-09

NUMBER OF SEQ ID NOS: 45

SOFTWARE: Wordperfect 5.1 (saved in ASCII format)

SEQ ID NO 13

LENGTH: 295

TYPE: PRT

ORGANISM: Mus musculus

FEATURE:

OTHER INFORMATION: estrogen sulfotransferase

US-09-150-141-13

Query Match 15.6%; Score 50; DB 3; Length 295;

Best Local Similarity 88.9%; Pred. No. 0.28;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TYPKSGTGW 9

DB 45 TYPKSGTGW 53

GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.545
; CURRENT APPLICATION NUMBER: US/09/374,492
; EARLIER FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,141
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
; US-09-374-492-13

Query Match 15.6%; Score 50; DB 4; Length 295;
Best Local Similarity 88.9%; Pred. No. 0.28;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TYPKSGTXW 9
Db 45 TYPKSGTTW 53

RESULT 14
US-09-609-816-4
; Sequence 4, Application US/09609816
; Patent No. 6436684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 304
; TYPE: PRT
; ORGANISM: HUMAN
; US-09-609-816-4

Query Match 15.6%; Score 50; DB 4; Length 304;
Best Local Similarity 88.9%; Pred. No. 0.29;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TYPKSGTXW 9
Db 53 TYPKSGTTW 61

RESULT 15
US-09-609-816-5
; Sequence 5, Application US/09609816

Patent No. 6436684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 304
; TYPE: PRT
; ORGANISM: HUMAN
; US-09-609-816-5

Query Match 15.6%; Score 50; DB 4; Length 304;
Best Local Similarity 88.9%; Pred. No. 0.29;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TYPKSGTXW 9
Db 53 TYPKSGTTW 61

Search completed: June 5, 2003, 12:32:42
Job time : 28 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:31:28 ; Search time 46 Seconds
(without alignments)
520.690 Million cell updates/sec

Title: US-09-854-122-1
Perfect score: 320
Sequence: 1 TYPKSTWXXXXXXXXXXXXX.....XXXXXXXXXGKXGDKWXXFT 232

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US08 NEW PUB.pap.*
2: /cgn2_6/ptodata/1/pubpaa/PCT NEW PUB.pap.*
3: /cgn2_6/ptodata/1/pubpaa/US06 NEW PUB.pap.*
4: /cgn2_6/ptodata/1/pubpaa/US06 PUBCOMB.pap.*
5: /cgn2_6/ptodata/1/pubpaa/US07 NEW PUB.pap.*
6: /cgn2_6/ptodata/1/pubpaa/US07 PUBCOMB.pap.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS PUBCOMB.pap.*
8: /cgn2_6/ptodata/1/pubpaa/US08 PUBCOMB.pap.*
9: /cgn2_6/ptodata/1/pubpaa/US09 NEW PUB.pap.*
10: /cgn2_6/ptodata/1/pubpaa/US09 PUBCOMB.pap.*
11: /cgn2_6/ptodata/1/pubpaa/US10 NEW PUB.pap.*
12: /cgn2_6/ptodata/1/pubpaa/US10 PUBCOMB.pap.*
13: /cgn2_6/ptodata/1/pubpaa/US60 NEW PUB.pap.*
14: /cgn2_6/ptodata/1/pubpaa/US60 PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	16.9	295	9	US-10-206-839-3
2	54	16.9	295	10	US-09-854-122-21
3	53	16.6	296	9	US-09-981-353-101
4	51	15.9	232	10	US-09-854-122-1
5	50	15.6	283	9	US-10-199-330-8
6	50	15.6	283	9	US-10-199-330-9
7	50	15.6	283	9	US-10-199-330-10
8	50	15.6	283	9	US-10-199-334-8
9	50	15.6	283	9	US-10-199-334-9
10	50	15.6	283	9	US-10-199-334-10
11	50	15.6	283	9	US-09-839-446-40
12	50	15.6	283	10	US-09-898-570-40
13	50	15.6	304	9	US-10-199-330-4
14	50	15.6	304	9	US-10-199-330-5
15	50	15.6	304	9	US-10-199-330-6
16	50	15.6	304	9	US-10-199-334-4
17	50	15.6	304	9	US-10-199-334-5
18	50	15.6	304	9	US-10-199-334-6
19	50	15.6	305	9	US-09-839-446-26

20	50	15.6	305	10	US-09-898-570-26	Sequence 26, Appl
21	47	14.7	214	10	US-09-795-926-16	Sequence 16, Appl
22	47	14.7	265	10	US-09-795-926-6	Sequence 6, Appl
23	47	14.7	303	10	US-09-795-926-2	Sequence 2, Appl
24	46	14.4	320	10	US-09-854-122-20	Sequence 20, Appl
25	45	14.1	8	10	US-09-854-122-12	Sequence 12, Appl
26	42	13.1	9	10	US-09-854-122-13	Sequence 13, Appl
27	42	13.1	9	10	US-09-854-122-14	Sequence 14, Appl
28	41	12.8	302	10	US-09-854-122-18	Sequence 18, Appl
29	41	12.8	324	10	US-09-854-122-17	Sequence 17, Appl
30	39	12.2	331	10	US-09-854-122-16	Sequence 16, Appl
31	39	12.2	391	10	US-09-795-926-4	Sequence 4, Appl
32	36	11.2	110	10	US-09-795-926-8	Sequence 8, Appl
33	36	11.2	148	10	US-09-795-926-10	Sequence 10, Appl
34	36	11.2	148	10	US-09-795-926-12	Sequence 12, Appl
35	36	11.2	186	10	US-09-795-926-12	Sequence 22, Appl
36	35	10.9	42	9	US-10-004-530A-22	Sequence 11, Appl
37	35	10.9	42	12	US-10-003-674A-11	Sequence 12, Appl
38	35	10.9	42	12	US-10-003-674A-12	Sequence 4369, Ap
39	34	10.6	92	9	US-09-764-891-4369	Sequence 719, App
40	34	10.6	134	10	US-09-925-297-719	Sequence 1023, Ap
41	34	10.6	142	9	US-09-764-868-1023	Sequence 429, App
42	34	10.6	142	9	US-10-103-313-429	Sequence 610, App
43	34	10.6	241	9	US-10-102-806-610	Sequence 36, Appl
44	34	10.6	275	10	US-09-733-387-36	Sequence 16, Appl
45	34	10.6	294	10	US-09-733-387-16	

ALIGNMENTS

RESULT 1
US-10-206-839-3
; Sequence 3, Application US/10206839
; Publication No. US2003009977A1
; GENERAL INFORMATION:
; APPLICANT: Guida, Marco
; APPLICANT: Kurth, Janice
; TITLE OF INVENTION: Genotyping Human Phenol Sulfotransferase
; FILE REFERENCE: 4389-6 (formerly SEQ-16P)
; CURRENT APPLICATION NUMBER: US/10/206,839
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 09/328,174
; PRIOR FILING DATE: 1999-06-08
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 295
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-206-839-3

Query Match 16.9%; Score 54; DB 9; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.052;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 220 RKXGXDWXXFT 232
||| |||||
Db 257 RKGAGDWKTFT 269

RESULT 2
US-09-854-122-21
; Sequence 21, Application US/09854122
; Patent No. US2002001690A1
; GENERAL INFORMATION:
; APPLICANT: ALBERTE, RANDALL S.
; APPLICANT: SMITH, ROBERT
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA
; FILE REFERENCE: PHA-007.01
; CURRENT APPLICATION NUMBER: US/09/854,122
; CURRENT FILING DATE: 2001-09-10

; PRIOR APPLICATION NUMBER: 60/202,529
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 21
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-122-21

Query Match 16.9%; Score 54; DB 10; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.052;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDKXXFT 232
||| ||||| ||
Db 257 RKGAGDKWKTFT 269

RESULT 3
US-09-981-353-101
; Sequence 101, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 101
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 4874364CD1
US-09-981-353-101

Query Match 16.6%; Score 53; DB 9; Length 296;
Best Local Similarity 69.2%; Pred. No. 0.079;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDKXXFT 232
||| ||||| ||
Db 258 RKGAGDKWKTFT 270

RESULT 4
US-09-854-122-1
; Sequence 1, Application US/09854122
; Patent No. US20020016980A1
; GENERAL INFORMATION:
; APPLICANT: ALBERTE, RANDALL S.
; APPLICANT: SMITH, ROBERT
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA
; FILE REFERENCE: PHA-007.01
; CURRENT APPLICATION NUMBER: US/09/854,122
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/202,529
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Consensus
; OTHER INFORMATION: sequence
; NAME/KEY: MOD_RES

; LOCATION: (8)
; OTHER INFORMATION: N or T
; NAME/KEY: MOD RES
; LOCATION: (10)-(219)
; OTHER INFORMATION: Any amino acid; this region may vary in length from
; OTHER INFORMATION: 190-210 amino acids
; NAME/KEY: MOD RES
; LOCATION: (223)-(224)
; OTHER INFORMATION: Any amino acid
; NAME/KEY: MOD RES
; LOCATION: (229)-(230)
; OTHER INFORMATION: Any amino acid
US-09-854-122-1

Query Match 15.9%; Score 51; DB 10; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 RKGXXGDKXXFT 232
||| ||||| |||||
Db 220 RKGXXGDKXXFT 232

RESULT 5
US-10-199-330-8
; Sequence 8, Application US/10199330
; Publication No. US20020182681A1
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: CL000669DIV-1
; CURRENT APPLICATION NUMBER: US/10/199,330
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-330-8

Query Match 15.6%; Score 50; DB 9; Length 283;
Best Local Similarity 69.2%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDKXXFT 232
||| ||||| ||
Db 245 RKGWPGDKWNYFT 257

RESULT 6
US-10-199-330-9
; Sequence 9, Application US/10199330
; Publication No. US20020182681A1
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: CL000669DIV-1
; CURRENT APPLICATION NUMBER: US/10/199,330
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816

```

; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-330-9

Query Match      15.6%; Score 50; DB 9; Length 283;
Best Local Similarity 69.2%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMKXXFT 232
DB 245 RKGMPGDKNYFT 257

RESULT 7
US-10-199-330-10
; Sequence 10, Application US/10199330
; Publication No. US20020182681A1
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669DIV-1
; CURRENT APPLICATION NUMBER: US/10/199,330
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-330-10

Query Match      15.6%; Score 50; DB 9; Length 283;
Best Local Similarity 69.2%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMKXXFT 232
DB 245 RKGMPGDKNYFT 257

RESULT 8
US-10-199-334-8
; Sequence 8, Application US/10199334
; Publication No. US20030008354A1
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669DIV-3
; CURRENT APPLICATION NUMBER: US/10/199,334
; CURRENT FILING DATE: 2002-07-22

```

```

; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-334-8

Query Match      15.6%; Score 50; DB 9; Length 283;
Best Local Similarity 69.2%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMKXXFT 232
DB 245 RKGMPGDKNYFT 257

RESULT 9
US-10-199-334-9
; Sequence 9, Application US/10199334
; Publication No. US20030008354A1
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669DIV-3
; CURRENT APPLICATION NUMBER: US/10/199,334
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-334-9

Query Match      15.6%; Score 50; DB 9; Length 283;
Best Local Similarity 69.2%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMKXXFT 232
DB 245 RKGMPGDKNYFT 257

RESULT 10
US-10-199-334-10
; Sequence 10, Application US/10199334
; Publication No. US20030008354A1
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669DIV-3
; CURRENT APPLICATION NUMBER: US/10/199,334

```

```
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-334-10

Query Match      15.6%; Score 50; DB 9; Length 283;
Best Local Similarity 69.2%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      220 RKGXXGDKWXXFT 232
Db      245 RKGMPGDWKNYFT 257

RESULT 11
US-09-839-446-40
; Sequence 40, Application US/09839446
; Publication No. US2003050232A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: ELLERMAN, KAREN
; APPLICANT: MACDOUGALL, JOHN R.
; APPLICANT: SMITHSON, GLENDA
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; FILE REFERENCE: 15966-776
; CURRENT APPLICATION NUMBER: US/09/839,446
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,814
; PRIOR FILING DATE: 2001-02-27
; CURRENT APPLICATION NUMBER: US/09/839,446
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/218,591
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,814
; PRIOR FILING DATE: 2001-02-27
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-839-446-40

Query Match      15.6%; Score 50; DB 9; Length 283;
Best Local Similarity 69.2%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      220 RKGXXGDKWXXFT 232
Db      245 RKGMPGDWKNYFT 257

RESULT 12
US-09-898-570-40
; Sequence 40, Application US/09898570
; Patent No. US20020123612A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: ELLERMAN, KAREN
; APPLICANT: MACDOUGALL, JOHN R.
; APPLICANT: SMITHSON, GLENDA
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; FILE REFERENCE: 15966-776CIP
; CURRENT APPLICATION NUMBER: US/09/898,570
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,814
; PRIOR FILING DATE: 2001-02-27
; CURRENT APPLICATION NUMBER: US/10/199,330
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27

Query Match      15.6%; Score 50; DB 10; Length 283;
Best Local Similarity 69.2%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      220 RKGXXGDKWXXFT 232
Db      245 RKGMPGDWKNYFT 257

RESULT 13
US-10-199-330-4
; Sequence 4, Application US/10199330
; Publication No. US20020182681A1
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO00669DIV-1
; CURRENT APPLICATION NUMBER: US/10/199,330
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27
```

;; PRIOR APPLICATION NUMBER: 60/192,408
;; PRIOR FILING DATE: 2000-03-27
;; PRIOR APPLICATION NUMBER: 60/212,725
;; PRIOR FILING DATE: 2000-06-20
;; PRIOR APPLICATION NUMBER: 09/609,816
;; PRIOR FILING DATE: 2000-07-03
;; NUMBER OF SEQ ID NOS: 23
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4
;; LENGTH: 304
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-199-330-4

Query Match 15.6%; Score 50; DB 9; Length 304;
Best Local Similarity 88.9%; Pred. No. 0.28;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TYPKSGTXW 9
Db 53 TYPKSGTTW 61

RESULT 14
US-10-199-330-5
;; Sequence 5, Application US/10199330
;; Publication No. US20020182681A1
;; GENERAL INFORMATION:
;; APPLICANT: WOODAGE, Trevor et al.
;; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
;; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
;; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
;; FILE REFERENCE: CL000669DIV-1
;; CURRENT APPLICATION NUMBER: US/10/199,330
;; PRIOR FILING DATE: 2002-07-22
;; PRIOR APPLICATION NUMBER: 09/609,816
;; PRIOR FILING DATE: 2001-03-27
;; PRIOR APPLICATION NUMBER: 60/192,408
;; PRIOR FILING DATE: 2000-03-27
;; PRIOR APPLICATION NUMBER: 60/212,725
;; PRIOR FILING DATE: 2000-06-20
;; PRIOR APPLICATION NUMBER: 09/609,816
;; PRIOR FILING DATE: 2000-07-03
;; NUMBER OF SEQ ID NOS: 23
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 5
;; LENGTH: 304
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-199-330-5

Query Match 15.6%; Score 50; DB 9; Length 304;
Best Local Similarity 88.9%; Pred. No. 0.28;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TYPKSGTXW 9
Db 53 TYPKSGTTW 61

RESULT 15
US-10-199-330-6
;; Sequence 6, Application US/10199330
;; Publication No. US20020182681A1
;; GENERAL INFORMATION:
;; APPLICANT: WOODAGE, Trevor et al.
;; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
;; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
;; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
;; FILE REFERENCE: CL000669DIV-1
;; CURRENT APPLICATION NUMBER: US/10/199,330
;; CURRENT FILING DATE: 2002-07-22
;; PRIOR APPLICATION NUMBER: 09/609,816

;; PRIOR FILING DATE: 2001-03-27
;; PRIOR APPLICATION NUMBER: 60/192,408
;; PRIOR FILING DATE: 2000-03-27
;; PRIOR APPLICATION NUMBER: 60/212,725
;; PRIOR FILING DATE: 2000-06-20
;; PRIOR APPLICATION NUMBER: 09/609,816
;; PRIOR FILING DATE: 2000-07-03
;; NUMBER OF SEQ ID NOS: 23
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 6
;; LENGTH: 304
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-199-330-6

Query Match 15.6%; Score 50; DB 9; Length 304;
Best Local Similarity 88.9%; Pred. No. 0.28;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TYPKSGTXW 9
Db 53 TYPKSGTTW 61

Search completed: June 5, 2003, 12:40:33
Job time : 47 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:27:12 ; Search time 40 Seconds
(without alignments)
557.580 Million cell updates/sec

Title: US-09-854-122-1

Perfect score: 320

Sequence: 1 TYPRSGTWXXXXXXXXXXXXX.....XXXXXXXXXGDKWXXFT 232

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	16.9	291	2 S10329	aryl sulfotransfer
2	54	16.9	294	2 JC5000	aryl sulfotransfer
3	54	16.9	295	2 JC2523	aryl sulfotransfer
4	54	16.9	295	2 S52399	aryl sulfotransfer
5	54	16.9	295	2 S52791	aryl sulfotransfer
6	54	16.9	295	2 I57945	phenol-sulfating p
7	54	16.9	295	2 JC5248	aryl sulfotransfer
8	54	16.9	295	2 JC5249	aryl sulfotransfer
9	54	16.9	295	2 G01843	aryl sulfotransfer
10	54	16.9	295	2 G02924	phenol sulfotransf
11	54	16.9	295	2 A55451	aryl sulfotransfer
12	54	16.9	295	2 S52794	aryl sulfotransfer
13	53	16.6	282	2 A26136	sensence marker
14	53	16.6	285	2 T10086	alcohol sulfotrans
15	53	16.6	296	2 JC5885	thyroid hormone su
16	53	16.6	298	2 S28183	aryl sulfotransfer
17	52	16.2	249	2 E84072	hypothetical prote
18	52	16.2	284	2 A33569	alcohol sulfotrans
19	52	16.2	284	2 I52849	alcohol sulfotrans
20	52	16.2	285	1 I38548	alcohol sulfotrans
21	52	16.2	287	2 JE0152	alcohol sulfotrans
22	52	16.2	286	2 A54026	alcohol sulfotrans
23	52	16.2	294	2 JC2229	estrogen sulfotran
24	52	16.2	295	2 JE0197	phenol sulfotransf
25	52	16.2	296	2 A44011	adrenocortical est
26	52	16.2	338	2 JC7282	hydroxyarylamine s
27	52	16.2	338	2 JE0196	hydroxysteroid sul
28	51	15.9	287	2 JC4531	alcohol sulfotrans
29	51	15.9	295	2 S29045	estrone sulfotrans

30 51 15.9 296 2 JC7283 hydroxyarylamine s
31 51 15.9 299 2 JC5884 thyroid hormone su
32 51 15.9 299 2 JE0186 sulfotransferase (
33 51 15.9 311 2 B40216 flavonol 3'-sulfot
34 50 15.6 295 2 I53296 testis-specific es
35 50 15.6 304 2 A49098 N-hydroxyarylamine
36 49 15.3 284 2 I60190 alcohol sulfotrans
37 49 15.3 285 2 I65760 alcohol sulfotrans
38 49 15.3 301 2 JW0078 amine sulfotransfe
39 46 14.4 295 2 A1930 estrone sulfotrans
40 46 14.4 295 2 I56606 estrogen sulfotran
41 46 14.4 295 2 I73679 estrogen sulfotran
42 46 14.4 320 2 A40216 flavonol 4'-sulfot
43 46 14.4 333 2 A84523 probable steroid s
44 45 14.1 314 2 T06012 hypothetical prote
45 45 14.1 323 2 T47447 sulfotransferase-1

ALIGNMENTS

RESULT 1

S10329

aryl sulfotransferase (EC 2.8.2.1) IV - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 18-Jun-1999

C:Accession: S10329; A44932

R:Ozawa, S.; Nagata, K.; Gong, D.; Yamazoe, Y.; Kato, R.

Nucleic Acids Res. 18, 4001, 1990

A:Title: Nucleotide sequence of a full-length cDNA (PST-1) for aryl sulfotransferase from

A:Reference number: S10329; MUID:90326537; PMID:2374726

A:Accession: S10329

A:Molecule type: mRNA

A:Residues: 1-291 <OZA>

A:Cross-references: EMBL:X52883; NID:G55760; PIDN:CAA37065.1; PID:G55761

R:Yerokun, T.; Etheredge, J.L.; Norton, T.R.; Carter, H.A.; Chung, K.H.; Birckbichler, P.

Cancer Res. 52, 4779-4786, 1992

A:Title: Characterization of a complementary DNA for rat liver aryl sulfotransferase IV

used hepatocarcinogenesis.

A:Reference number: A44932; MUID:92379794; PMID:1511441

A:Accession: A44932

A>Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 7-291 <VER>

A:Cross-references: GB:X68640; GB:S42994; NID:G55764; PIDN:CAA48604.1; PID:G55765

A:Experimental source: liver

A>Note: sequence extracted from NCBI backbone (NCBIN:112171, NCBIIP:112178)

C:Superfamily: alcohol sulfotransferase

C:Keywords: sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 291;

Best Local Similarity 69.2%; Pred. No. 0.017; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0;

QY 220 RKGXXGDKWXXFT 232

Db 253 RKGTTGDKWXXFT 265

RESULT 2

JC5000

aryl sulfotransferase (EC 2.8.2.1), phenol-preferring - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 13-Dec-1996 #sequence_revision 21-Jan-1997 #text_change 20-Apr-2000

C:Accession: JC5000; S59269; S59070

R:Henry, T.; Kliewer, B.; Palmatier, R.; Ulphani, J.S.; Beckmann, J.D.

Gene 174, 221-224, 1996

A:Title: Isolation and characterization of a bovine gene encoding phenol sulfotransferase

A:Reference number: JC5000; MUID:97045816; PMID:8890738

A:Accession: JC5000

A:Molecule type: DNA

A:Residues: 1-294 <HEN>

A:Cross-references: GB:U34753; NID:G1276632; PIDN:AAC48677.1; PID:G1381041

R;Schauss, S.J.; Henry, T.; Palmatier, R.; Halvorson, L.; Dannenbring, R.; Beckmann, J.D.
Biochem. J. 311, 209-217, 1995
A;Title: Characterization of bovine tracheobronchial phenol sulphotransferase cDNA and c
A;Reference number: S59070; MUID:96003918; PMID:7575456
A;Accession: S59269
A;Molecule type: mRNA
A;Residues: 1-294 <SCH>
A;Cross-references: GB:U35253; NID:g1006836; PIDN:AAA85510.1; PID:g1006837
A;Accession: S59070
A;Molecule type: protein
A;Residues: 23-36738-47;79-96;134-144;207-226;234-248 <SCW>
C;Comment: This enzyme belongs to a family of cytosolic enzymes that catalyze the trans
C;Genetics:
A;Gene: pst
A;Introns: 50/1; 93/3; 124/3; 167/1; 198/3; 259/1
C;Superfamily: alcohol sulfotransferase
C;Keywords: sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 294;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXGDKWXXFT 232
Db 257 RKGITGDWKTFT 269

RESULT 3
JC2523
aryl sulfotransferase (EC 2.8.2.1) brain isoform - human
N;Alternate names: phenol sulfotransferase
C;Species: Homo sapiens (man)
C;Date: 19-May-1995 #sequence_revision 14-Jul-1995 #text_change 28-May-1999
C;Accession: JC2523
R;Hwang, S.R.; Kohn, A.B.; Hook, V.Y.H.
Biochem. Biophys. Res. Commun. 207, 701-707, 1995
A;Title: Molecular cloning of an isoform of phenol sulfotransferase from human brain hip
A;Reference number: JC2523; MUID:95169114; PMID:7864863
A;Accession: JC2523
A;Molecule type: mRNA
A;Residues: 1-295 <HWA>
A;Cross-references: GB:U26309; NID:g847762; PIDN:AAA67895.1; PID:g847763
C;Experimental source: hippocampus
C;Superfamily: alcohol sulfotransferase
C;Keywords: sulfotransferase
F;70/Active site: Cys #status predicted

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXGDKWXXFT 232
Db 257 RKGWAGDKWKTFT 269

RESULT 4
S52399
aryl sulfotransferase (EC 2.8.2.1) - human
C;Species: Homo sapiens (man)
C;Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 18-Jun-1999
C;Accession: S52399
R;Jones, A.L.; Hagen, M.; Coughtrie, M.W.H.; Roberts, R.C.; Glatt, H.R.
Submitted to the EMBL Data Library, February 1995
A;Description: Human platelet phenol sulfotransferases: cDNA cloning, stable expression
A;Reference number: S52398
A;Accession: S52399
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-295 <JON>
A;Cross-references: EMBL:X84654; NID:g671641; PIDN:CAA59147.1; PID:g671642
C;Superfamily: alcohol sulfotransferase
C;Keywords: sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXGDKWXXFT 232
Db 257 RKGWAGDKWKTFT 269

RESULT 5
S52791
aryl sulfotransferase (EC 2.8.2.1) - human
C;Species: Homo sapiens (man)
C;Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 18-Jun-1999
C;Accession: S52791
R;Ozawa, S.; Nagata, K.; Shimada, M.; Ueda, M.; Tazuke, T.; Yamazoe, Y.; Kato, R.
Submitted to the EMBL Data Library, March 1994
A;Description: Primary structures and properties of two related aryl sulfotransferases
A;Reference number: S52791
A;Accession: S52791
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-295 <OZA>
A;Cross-references: EMBL:X78282; NID:g758594; PIDN:CAA55088.1; PID:g758595
C;Superfamily: alcohol sulfotransferase
C;Keywords: sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXGDKWXXFT 232
Db 257 RKGWAGDKWKTFT 269

RESULT 6
I57945
phenol-sulfating phenol sulfotransferase - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: I57945
R;Wilborn, T.W.; Comer, K.A.; Dooley, T.P.; Reardon, I.M.; Heinrikson, R.L.; Falany, C.N.
Mol. Pharmacol. 43, 70-77, 1993
A;Title: Sequence analysis and expression of the cDNA for the phenol-sulfating form of h
A;Reference number: I57945; MUID:93140712; PMID:8423770
A;Accession: I57945
A;Status: preliminary; translated from GB/EMBL/DBBJ
A;Molecule type: mRNA
A;Residues: 1-295 <RES>
A;Cross-references: GB:L19999; NID:g307342; PIDN:AAA99892.1; PID:g307343
C;Superfamily: alcohol sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXGDKWXXFT 232
Db 257 RKGWAGDKWKTFT 269

RESULT 7
JC5248
aryl sulfotransferase (EC 2.8.2.1) HAST2 - human
N;Alternate names: aryl sulfotransferase 1; phenol sulfotransferase; sulfokinase
C;Species: Homo sapiens (man)
C;Date: 13-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 18-Jun-1999
C;Accession: JC5248; JN0714; JN0479; G01421
R;Dooley, T.P.; Huang, Z.
Biochem. Biophys. Res. Commun. 228, 134-140, 1996
A;Title: Genomic organization and DNA sequences of two human phenol sulfotransferase gene

A;Reference number: JC5248; MUID:97069665; PMID:8912648
A;Accession: JC5248
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-295 <DOO>
A;Cross-references: GB:U71086; NID:91575787; PIDN:AAB09597.1; PID:91575788
R;Zhu, X.; Veronese, M.E.; Bernard, C.C.A.; Sansom, L.N.; McManus, M.E.
Biochem. Biophys. Res. Commun. 195, 120-127, 1993
A;Title: Identification of two human brain aryl sulfotransferase cDNAs.
A;Reference number: JN0713; MUID:93371391; PMID:8363592
A;Accession: JN0714
A;Molecule type: mRNA
A;Residues: 1-295 <ZHU>
A;Cross-references: GB:U119955; NID:9306454; PIDN:AAA02935.1; PID:9306455
R;Zhu, X.; Veronese, M.E.; Sansom, L.N.; McManus, M.E.
Biochem. Biophys. Res. Commun. 192, 671-676, 1993
A;Title: Molecular characterization of a human aryl sulfotransferase cDNA.
A;Reference number: JN0479; MUID:93249441; PMID:8484775
A;Accession: JN0479
A;Molecule type: mRNA
A;Residues: 1-295 <ZH2>
A;Cross-references: GB:U01819
A;Experimental source: liver
A;Note: at least two sizes of aryl sulfotransferase mRNA are detected in human liver and submitted to the EMBL Data Library, April 1994
A;Reference number: G06944
A;Accession: G01421
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-295 <ZH3>
A;Cross-references: EMBL:U09031; NID:9495486; PIDN:AAA18613.1; PID:9495487
C;Comment: Sulfotransferases catalyze the transfer of the sulfonate group from 3'-phosphoric aromatic amines, N-hydroxylated aromatic amines and steroids. This results usually in the aromatic amines, N-hydroxylated aromatic amines and steroids.
C;Comment: This sulfotransferase acts preferentially on various phenolic substrates.
C;Genetics:
A;Gene: GDB:STP1
A;Cross-references: GDB:138288; OMIM:171150
A;Map position: 16p12.1-16p12.1
C;Superfamily: alcohol sulfotransferase
C;Keywords: sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDWKKXFT 232
||| ||||| ||
Db 257 RKGWAGDWKTTFT 269

RESULT 8
JC5249
aryl sulfotransferase (EC 2.8.2.1) 2 - human
A;Alternate names: phenol sulfotransferase; sulfokinase
C;Species: Homo sapiens (man)
C;Date: 13-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 18-Jun-1999
C;Accession: JC5249
R;Dooley, T.P.; Huang, Z.
Biochem. Biophys. Res. Commun. 228, 134-140, 1996
A;Title: Genomic organization and DNA sequences of two human phenol sulfotransferase genes
A;Reference number: JC5248; MUID:97069665; PMID:8912648
A;Accession: JC5249
A;Molecule type: mRNA
A;Residues: 1-295 <DOO>
A;Cross-references: GB:U76619; NID:91666920; PIDN:AAB18753.1; PID:91666921
C;Comment: This enzyme catalyzes the transfer of the sulfonate group from phosphoadenosine.
C;Genetics:
A;Gene: stp2
A;Map position: 16p12.1-pl1.2
C;Superfamily: alcohol sulfotransferase
C;Keywords: sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDWKKXFT 232
||| ||||| ||
Db 257 RKGWAGDWKTTFT 269

RESULT 9
G01843
aryl sulfotransferase - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 18-Jun-1999
C;Accession: G01843
R;Zhu, X.
submitted to the EMBL Data Library, May 1995
A;Reference number: G08606
A;Accession: G01843
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-295 <ZHU>
A;Cross-references: EMBL:U28169; NID:9881502; PIDN:AAB09658.1; PID:9881503
C;Superfamily: alcohol sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDWKKXFT 232
||| ||||| ||
Db 257 RKGWAGDWKTTFT 269

RESULT 10
G02924
phenol sulfotransferase subunit - crab-eating macaque (fragment)
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 18-Jun-1999
C;Accession: G02924
R;Ogura, K.
submitted to GenBank, May 1996
A;Reference number: H01975
A;Accession: G02924
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-295 <OGU>
A;Cross-references: GB:D85514; NID:91339922; PIDN:BAA12822.1; PID:91339923
C;Superfamily: alcohol sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDWKKXFT 232
||| ||||| ||
Db 257 RKGWAGDWKTTFT 269

RESULT 11
A55451
aryl sulfotransferase (EC 2.8.2.1) HAST3 / estrogen sulfotransferase EST - human
N;Alternate names: gene STM protein; thermolabile phenol sulfotransferase
C;Species: Homo sapiens (man)
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 18-Jun-1999
C;Accession: A55451; JN0713; JC2067; S52398; I38551; I57679; I52299; PC2031
R;Bernier, F.; Leblanc, G.; Labrie, F.; Luu-The, V.
J. Biol. Chem. 269, 28200-28205, 1994
A;Title: Structure of human estrogen and aryl sulfotransferase gene. Two mRNA species isolated.
A;Reference number: A55451; MUID:95050600; PMID:7961757
A;Accession: A55451
A;Molecule type: DNA
A;Residues: 1-295 <BER>

A/Cross-references: GB:L34160
R/Zhu, X.; Veronese, M.E.; Bernard, C.C.A.; Sansom, L.N.; McManus, M.E.
Biochem. Biophys. Res. Commun. 195, 120-127, 1993
A/Title: Identification of two human brain aryl sulfotransferase cDNAs.
A/Reference number: JN0713; MUID:93371391; PMID:8363592
A/Accession: JN0713
A/Molecule type: mRNA
A/Residues: 1-295 <ZHU>
A/Cross-references: GB:L19956; NID:g306456; PIDN:AAA02943.1; PID:g306457
A/Experimental source: brain
R/Wood, T.C.; Aksoy, I.A.; Aksoy, S.; Weinshilboum, R.M.
Biochem. Biophys. Res. Commun. 198, 1119-1127, 1994
A/Title: Human liver thermolabile phenol sulfotransferase: cDNA cloning, expression and
A/Reference number: JC2067; MUID:94161723; PMID:8117269
A/Accession: JC2067
A/Molecule type: mRNA
A/Residues: 1-295 <WO>
A/Cross-references: GB:U08032; NID:g468256; PIDN:AAA17723.1; PID:g468257
A/Experimental source: liver
R/Jones, A.L.; Hagen, M.; Coughtrie, M.W.H.; Roberts, R.C.; Glatt, H.R.
submitted to the EMBL Data Library, February 1995
A/Description: Human platelet phenolsulfotransferases: cDNA cloning, stable expression and
A/Reference number: S52398
A/Accession: S52398
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-295 <JON>
A/Cross-references: EMBL:X84653; NID:g671533; PIDN:CAA59146.1; PID:g671534
R/Aksoy, I.A.; Callen, D.P.; Apostolou, S.; Her, C.; Weinshilboum, R.M.
Genomics 23, 275-277, 1994
A/Title: Thermolabile phenol sulfotransferase gene (STM): localization to human chromosome
A/Reference number: A55725; MUID:95130098; PMID:7829089
A/Accession: I38551
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 138-198 <RES>
A/Cross-references: EMBL:U08099; NID:g488284; PIDN:AAA82126.1; PID:g488285
R/Bernier, F.; Lopez-Solache, I.; Labrie, F.; Luu-The, V.
Mol. Cell. Endocrinol. 99, 11-15, 1994
A/Title: Cloning and expression of cDNA encoding human placental estrogen sulfotransferase
A/Reference number: I57679; MUID:94244847; PMID:8187952
A/Accession: I57679
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-295 <RE2>
A/Cross-references: GB:L25275; NID:g463124; PIDN:AAA36523.1; PID:g463125
R/Dooley, T.P.; Probst, P.; Munroe, P.B.; Mole, S.E.; Liu, Z.; Doggett, N.A.
Biochem. Biophys. Res. Commun. 205, 1325-1332, 1994
A/Title: Genomic organization and DNA sequence of the human catecholamine-sulfating phenol
A/Reference number: I52299; MUID:95100963; PMID:7802665
A/Accession: I52299
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-295 <RE3>
A/Cross-references: GB:S74971; NID:g833898
C/Comment: This enzyme transfers a sulfonate moiety from 3'-phosphoadenosine-5-phosphosulphate
C/Comment: Brain aryl sulfotransferase and placental estrogen sulfotransferase are transcribed
C/Comment: differ only in their 5'-untranslation regions.
C/Genetics:
A/Gene: GDB:STP
A/Cross-references: GDB:342104; OMIM:600043
A/Map position: 16p11.2-16p11.2
A/Introns: 50/1; 92/1; 124/3; 167/1; 198/3; 259/1
A/Suprafamily: alcohol sulfotransferase
C/Keywords: brain; placenta; sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMKXXFT 232
||| |||||
Db 257 RKGMDGDKTFT 269

RESULT 12

S52794
aryl sulfotransferase (EC 2.8.2.1) - human
C/Species: Homo sapiens (man)
C/Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 18-Jun-1999
C/Accession: S52794
R/Ozawa, S.; Nagata, K.; Shimada, M.; Ueda, M.; Teuzuki, T.; Yamazoe, Y.; Kato, R.
submitted to the EMBL Data Library, March 1994
A/Description: Primary structures and properties of two related aryl sulfotransferases;
A/Reference number: S52791
A/Accession: S52791
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-295 <OZA>
A/Cross-references: EMBL:X78283; NID:g758596; PIDN:CAA55089.1; PID:g758597
C/Suprafamily: alcohol sulfotransferase
C/Keywords: sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMKXXFT 232
||| |||||
Db 257 RKGMDGDKTFT 269

RESULT 13

A26136
senescence marker protein 2, hepatic - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 07-Feb-1997
C/Accession: A26136
R/Chatterjee, B.; Majumdar, D.; Ozbilien, O.; Murty, C.V.R.; Roy, A.K.
J. Biol. Chem. 262, 822-825, 1987
A/Title: Molecular cloning and characterization of cDNA for androgen-repressible rat liver
A/Reference number: A26136; MUID:87109186; PMID:3805009
A/Accession: A26136
A/Molecule type: mRNA
A/Residues: 1-282 <CHA>
C/Suprafamily: alcohol sulfotransferase
C/Keywords: liver

Query Match 16.6%; Score 53; DB 2; Length 282;
Best Local Similarity 69.2%; Pred. No. 0.026;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMKXXFT 232
||| |||||
Db 244 RKGTTGDMKXHT 256

RESULT 14

T10086
alcohol sulfotransferase (EC 2.8.2.2) - mouse
C/Species: Mus musculus (house mouse)
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 26-Aug-1999
C/Accession: T10086
R/Kong, A.N.T.
submitted to the EMBL Data Library, June 1995
A/Description: cDNA cloning of the mouse liver alcohol/hydroxysteroid form of (mSta2) sul
A/Reference number: Z16940
A/Accession: T10086
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-285 <KON>
A/Cross-references: EMBL:L27121; NID:g496151; PID:g496152
A/Experimental source: strain BALB/c, liver
C/Genetics:
A/Gene: STA2
C/Suprafamily: alcohol sulfotransferase

Query Match 16.9%; Score 54; DB 2; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.017;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDMKXXFT 232
||| |||||
Db 257 RKGMDGDKTFT 269

C;Keywords: sulfotransferase

Query Match 16.6%; Score 53; DB 2; Length 285;
Best Local Similarity 69.2%; Pred. No. 0.026;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDMKXXFT 232
Db 247 RKGTTGDMKNHFT 259

RESULT 15

JC5885
thyroid hormone sulfotransferase (EC 2.8.2.-) B2 - human
C;Species: Homo sapiens (man)
C;Date: 11-Mar-1998 #sequence_revision 11-Mar-1998 #text_change 20-Jun-2000
C;Accession: JC5885
R;Fujita, K.; Nagata, K.; Ozawa, S.; Sasano, H.; Yamazoe, Y.
J. Biochem. 122, 1052-1061, 1997
A;Title: Molecular cloning and characterization of rat ST1B1 and human ST1B2 cDNAs, encoding a novel sulfotransferase
A;Reference number: JC5884; MUID:98104061; PMID:9443824
A;Accession: JC5885
A;Molecule type: mRNA
A;Residues: 1-296 <FUG>
A;Cross-references: DDBJ:D89479; NID:G2826145; PIDN:BAA24547.1; PID:G2826146
A;Experimental source: liver
C;Comment: This enzyme mediates the transfer of SO3- from 3'-phosphoadenosine 5'-phosphate to various substrates
C;Superfamily: alcohol sulfotransferase
C;Keywords: sulfotransferase

Query Match 16.6%; Score 53; DB 2; Length 296;
Best Local Similarity 69.2%; Pred. No. 0.027;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXXGDMKXXFT 232
Db 258 RKGTTGDMKNHFT 270

Search completed: June 5, 2003, 12:32:08
Job time : 41 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:17:22 ; Search time 23 Seconds
(without alignments)

418.370 Million cell updates/sec

Title: US-09-854-122-1

Perfect score: 320

Sequence: 1 TYPKSGTXXWXXXXXXXXXXXXX.....XXXXXXXXXXGDXGDKXXFT 232

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	16.9	291	1 SUAR_RAT	P17988 rattus norv
2	54	16.9	294	1 SUPP_BOVIN	P50227 bos taurus
3	54	16.9	295	1 SUP1_HUMAN	P50225 homo sapien
4	54	16.9	295	1 SUP2_HUMAN	P50226 homo sapien
5	54	16.9	295	1 SUPM_HUMAN	P50224 h monoamine
6	54	16.9	295	1 SUPP_MACFA	P52846 macaca fasc
7	53	16.6	285	1 SUH2_MOUSE	P50236 mus musculu
8	53	16.6	285	1 SUH2_RAT	P07631 rattus norv
9	53	16.6	291	1 SUAR_MOUSE	P52840 mus musculu
10	52	16.2	283	1 SUHS_RAT	P15709 rattus norv
11	52	16.2	284	1 SUHA_HUMAN	Q06520 homo sapien
12	52	16.2	285	1 SUH1_MOUSE	P52843 mus musculu
13	52	16.2	286	1 SUHA_CAVPO	P50234 cavia porce
14	52	16.2	294	1 SUOE_HUMAN	P49888 homo sapien
15	52	16.2	296	1 SIC1_HUMAN	O00338 homo sapien
16	52	16.2	286	1 SIC1_RABIT	O46503 oryctolagus
17	52	16.2	296	1 STK1_RAT	P49887 rattus norv
18	52	16.2	296	1 SUOE_CAVPO	Q75897 homo sapien
19	52	16.2	302	1 SIC2_HUMAN	P52841 cavia porce
20	51	15.9	286	1 SUHB_CAVPO	P19217 bos taurus
21	51	15.9	295	1 SUOE_BOVIN	Q9WUW9 rattus norv
22	51	15.9	296	1 STK2_RAT	P52847 rattus norv
23	51	15.9	299	1 SUDY_RAT	P52838 flaveria bi
24	51	15.9	309	1 FSTL_FLABI	P52836 flaveria ch
25	51	15.9	311	1 F3ST_FLABI	P52835 flaveria bi
26	51	15.9	312	1 F3ST_FLABI	P49891 mus musculu
27	50	15.6	295	1 SUOT_MOUSE	P50237 rattus norv
28	50	15.6	304	1 SUAC_RAT	P22789 rattus norv
29	49	15.3	283	1 SUHA_RAT	P52842 macaca fasc
30	49	15.3	284	1 SUHA_MACFA	P50235 rattus norv
31	49	15.3	285	1 SUH3_RAT	P52844 rattus norv
32	46	14.4	295	1 SUO1_RAT	P52845 rattus norv
33	46	14.4	295	1 SUO2_RAT	P52845 rattus norv

34	46	14.4	295	1 SUO3_RAT	P49889 rattus norv
35	46	14.4	295	1 SUO6_RAT	P49890 rattus norv
36	46	14.4	320	1 F4ST_FLACH	P52837 flaveria ch
37	41	12.8	302	1 FSTL_ARATH	P52839 arabidopsis
38	39	12.2	284	1 S4A1_HUMAN	Q9B501 homo sapien
39	39	12.2	284	1 S4A1_MOUSE	Q9D597 mus musculu
40	36	11.2	350	1 UNRI_HUMAN	Q9Y3f4 homo sapien
41	36	11.2	351	1 UNRI_MOUSE	Q9Z1z2 mus musculu
42	36	11.2	3390	1 POLG_DEN3	P27915 d genome po
43	35	10.9	42	1 GIP_FIG	P01281 sus scrofa
44	35	10.9	144	1 GIP_RAT	Q06145 rattus norv
45	35	10.9	153	1 GIP_HUMAN	P09681 homo sapien

ALIGNMENTS

RESULT 1
SUAR_RAT
ID SUAR_RAT STANDARD; PRT; 291 AA.
AC F17988;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aryl sulfotransferase (EC 2.8.2.1) (Phenol sulfotransferase) (PST-1)
DE (Sulfokinase) (Aryl sulfotransferase IV) (ASRIV) (Tyrosine-ester
DE sulfotransferase) (Minoxidil sulfotransferase).
GN SULT1A1 OR ST1A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=90326537; PubMed=2374726;
RA Ozawa S., Nagata K., Gong D., Yamazoe Y., Kato R.;
RT "Nucleotide sequence of a full-length cDNA (PST-1) for aryl
RT sulfotransferase from rat liver.";
RL Nucleic Acids Res. 18:4001-4001(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=94131305; PubMed=8299966;
RA Khan A.S., Taylor B.R., Chung K., Etheredge J., Gonzales R.,
RA Ringer D.P.;
RT "Genomic structure of rat liver aryl sulfotransferase IV-encoding
RT gene.";
RL Gene 137:321-326(1993).
RN [3]
RP SEQUENCE OF 7-291 FROM N.A.
RX MEDLINE=92379794; PubMed=1511441;
RA Birckbichler P.J., Ringer D.P.;
RT "Characterization of a complementary DNA for rat liver aryl
RT sulfotransferase IV and use in evaluating the hepatic gene transcript
RT levels of rats at various stages of 2-acetylaminofluorene-induced
RT hepatocarcinogenesis.";
RL Cancer Res. 52:4779-4786(1992).
RN [4]
RP CHARACTERIZATION.
RC TISSUE=Liver;
RX MEDLINE=93191709; PubMed=8447833;
RA Cruickshank D., Sansom L.N., Veronese M.E., Mojarrahi B.,
RA McManus M.E., Zhu X.;
RT "cDNA expression studies of rat liver aryl sulphotransferase.";
RL Biochem. Biophys. Res. Commun. 191:295-301(1993).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=95196680; PubMed=7889867;
RA Yamazoe Y., Ozawa S., Nagata K., Gong D.-W., Kato R.;
RT "Characterization and expression of hepatic sulfotransferase involved
RT in the metabolism of N-substituted aryl compounds.";

RL Environ. Health Perspect. 102:99-103(1994).
 RN [6]
 RP SEQUENCE OF 63-68, AND CHARACTERIZATION.
 RX MEDLINE=95074030; PubMed=7982943;
 RA Zheng Y., Bergold A., Duffel M.W.;
 RT Affinity labeling of aryl sulfotransferase IV. Identification of a
 RT peptide sequence at the binding site for 3'-phosphoadenosine-5'-
 RT phosphosulfate.";
 RL J. Biol. Chem. 269:30313-30319(1994).
 RN [7]
 RP CHARACTERIZATION.
 RX MEDLINE=94306583; PubMed=8033271;
 RA Runge-Morris M.A.;
 RT Sulfotransferase gene expression in rat hepatic and extrahepatic
 RT tissues.";
 RL Chem. Biol. Interact. 92:67-76(1994).
 CC -1- FUNCTION: CATALYZES O-SULFATION OF PHENOLS, N-O- SULFATION OF
 CC MINOXIDIL AND TYROSINE ESTERS.
 CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + a phenol =
 CC adenosine 3',5'-bisphosphate + an aryl sulfate.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: LIVER, KIDNEY, HEART AND COLON.
 CC -1- INDUCTION: INDUCED BY ANDROGENS AND SUPPRESSED BY ESTROGENS. THE
 CC EXPRESSION IS UNDER THE INFLUENCE OF PITUITARY GROWTH HORMONE AND
 CC THYROID HORMONE.
 CC -1- PFM: THE N-TERMINUS IS BLOCKED.
 CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L19998; AAA1644.1; -;
 DR EMBL; X52883; CAA37065.1; -;
 DR EMBL; L16241; -; NOT ANNOTATED_CDS.
 DR EMBL; X68640; CAA48604.1; -;
 DR PIR; S10329; S10329.
 DR HSP; P50224; 1CJM.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 DR Transferrase; Steroid metabolism.
 FT BINDING 255 261 PAPS BINDING SITE (POTENTIAL).
 SQ SEQUENCE 291 AA; 33906 MW; 95C66C72923DB872 CRC64;
 Query Match 16.9%; Score 54; DB 1; Length 291;
 Best Local Similarity 69.2%; Pred. No. 0.0055;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 220 RKGGXGDKWXXFT 232
 DB 253 RKGTGDKWNTFT 265
 RESULT 2
 SUPP BOVIN
 ID SUPP BOVIN STANDARD; PRT; 294 AA.
 AC P50227;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Phenol-sulfating phenol sulfotransferase (EC 2.8.2.1) (P-PST).
 GN STP.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;

RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP TISSUE=Tracheobronchial;
 RX MEDLINE=96003918; PubMed=7575456;
 RA Schauss S.J., Henry T., Palmatier R., Halvorson L., Dannenbring R.,
 RA Beckmann J.D.;
 RT "Characterization of bovine tracheobronchial phenol sulphotransferase
 RT cDNA and detection of mRNA regulation by cortisol.";
 RL Biochem. J. 311:209-217(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Henry T., Klierer B., Palmatier R., Ulphani J., Beckmann J.D.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 24-57 FROM N.A.
 RA Noneman D.J., Shibuya H., Johnson G.S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE SULFATE CONJUGATION OF CATECHOLAMINES AND
 CC OF PHENOLIC DRUGS.
 CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + a phenol =
 CC adenosine 3',5'-bisphosphate + an aryl sulfate.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: DISTAL LUNG PARENCHYMA.
 CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U35253; AAA85510.1; -;
 DR EMBL; U34753; AAC48677.1; -;
 DR EMBL; L33828; AAA56789.1; -;
 DR HSP; P50224; 1CJM.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 DR Transferrase; Steroid metabolism.
 FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).
 SQ SEQUENCE 294 AA; 34017 MW; 8ADEE67D47E69737 CRC64;
 Query Match 16.9%; Score 54; DB 1; Length 294;
 Best Local Similarity 69.2%; Pred. No. 0.0055;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 220 RKGGXGDKWXXFT 232
 DB 257 RKGTGDKWSTFT 269
 RESULT 3
 SUP1 HUMAN
 ID SUP1 HUMAN STANDARD; PRT; 295 AA.
 AC P50225; Q92818; Q9BVU6; Q9UGG7;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phenol-sulfating phenol sulfotransferase 1 (EC 2.8.2.1) (P-PST)
 DE (Thermostable phenol sulfotransferase) (Tg-PST) (HASTI/HAST2) (STIA3).
 GN SUL1A1 OR STP1 OR STP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Liver;
 RX MEDLINE=93249441; PubMed=8484775;
 RA Zhu X., Veronese M.E., Sansom L.N., McManus M.E.;

"Molecular characterization of a human aryl sulfotransferase cDNA.";
 Biochem. Biophys. Res. Commun. 192:671-676(1993).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Zhu X., Veronese M.E., Bernard C.C., Sansom L.N., McManus M.E.;
 RL "Identification of two human brain aryl sulfotransferase cDNAs.";
 Biochem. Biophys. Res. Commun. 195:120-127(1993).
 [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Hippocampus;
 RA Huang S.-R., Kohn A.B., Hook V.Y.H.;
 RL "Molecular cloning of an isoform of phenol sulfotransferase from human brain hippocampus.";
 Biochem. Biophys. Res. Commun. 207:701-707(1995).
 [4]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Liver;
 RA Wilborn T.W., Comer K.A., Dooley T.P., Reardon I.M.,
 RL Heinrichson R.L., Falany C.N.;
 RL "Sequence analysis and expression of the cDNA for the phenol-sulfating form of human liver phenol sulfotransferase.";
 Mol. Pharmacol. 43:70-77(1993).
 [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RA Jones A.L., Hagen M., Coughtrie M.W., Roberts R.C., Glatt H.;
 RL "Human platelet phenol-sulfotransferases: cDNA cloning, stable expression in V79 cells and identification of a novel allelic variant of the phenol-sulfating form.";
 Biochem. Biophys. Res. Commun. 208:855-862(1995).
 [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Ozawa S., Nagata K., Shimada M., Ueda M., Tsuzuki T., Yamazoe Y.,
 RL Kato R.;
 RL "Primary structures and properties of two related forms of aryl sulfotransferases in human liver.";
 Pharmacogenetics 5:S135-S140(1995).
 [7]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94306556; PubMed=8033246;
 RA Yamazoe Y., Nagata K., Ozawa S., Kato R.;
 RL "Structural similarity and diversity of sulfotransferases.";
 Chem. Biol. Interact. 92:107-117(1994).
 [8]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96211162; PubMed=8924211;
 RA Bernier F., Soucy P., Iuu-The V.;
 RL "Human phenol sulfotransferase gene contains two alternative promoters: Structure and expression of the gene.";
 DNA Cell Biol. 15:367-375(1996).
 [9]
 RP SEQUENCE FROM N.A.
 RC Dooley T.P., Huang Z.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 [10]
 RP SEQUENCE FROM N.A.
 RA Raftogianis R.B., Her C., Weinshilboum R.M.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 [11]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Dajani R., Hood A.M., Coughtrie M.W.;
 RL "A single amino acid, Glu46, governs the substrate specificity of human dopamine sulfotransferase, SULT1A3.";
 Mol. Pharmacol. 54:942-948(1998).
 [12]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 [13]
 RP SEQUENCE OF 1-107 FROM N.A.
 RC MEDLINE=94117022; PubMed=8288252;
 RA Dooley T.P., Obermoller R.D., Leiter E.H., Chapman H.D.,
 RL Falany C.N., Deng Z., Siciliano M.J.;
 RL "Mapping of the phenol sulfotransferase gene (STP) to human chromosome 16p12.1-p11.2 and to mouse chromosome 7.";
 Genomics 18:440-443(1993).
 [14]
 RP CHARACTERIZATION.
 RC MEDLINE=94379981; PubMed=8093002;
 RA Veronese M.E., Burgess W., Zhu X., McManus M.E.;
 RL "Functional characterization of two human sulphotransferase cDNAs that encode monoamine- and phenol-sulphating forms of phenol sulphotransferase: substrate kinetics, thermal-stability and inhibitor-sensitivity studies.";
 Biochem. J. 302:497-502(1994).
 [15]
 RP MUTAGENESIS OF CYS-70.
 RC MEDLINE=94306582; PubMed=8033270;
 RA Falany C.N., Zhuang W., Falany J.L.;
 RL "Characterization of expressed human phenol-sulfating phenol sulfotransferase: effect of mutating cys70 on activity and thermostability.";
 Chem. Biol. Interact. 92:57-66(1994).
 [16]
 RP VARIANTS GLN-37; HIS-213 AND VAL-223.
 RC MEDLINE=98005125; PubMed=9345314;
 RA Raftogianis R.B., Wood T.C., Otterness D.M., Van Loon J.A.,
 RL Weinshilboum R.M.;
 RL "Phenol sulfotransferase pharmacogenetics in humans: association of common SULT1A1 alleles with TS PST phenotype.";
 Biochem. Biophys. Res. Commun. 239:298-304(1997).
 [17]
 RP VARIANT HIS-213.
 RC MEDLINE=20222641; PubMed=10762004;
 RA Engelke C.E., Meinel W., Boeing H., Glatt H.;
 RL "Association between functional genetic polymorphisms of human sulfotransferases 1A1 and 1A2.";
 Pharmacogenetics 10:163-169(2000).
 CC -!- FUNCTION: CATALYZES THE SULFATE CONJUGATION OF CATECHOLAMINES, PHENOLIC DRUGS AND NEUROTRANSMITTERS. IS ALSO RESPONSIBLE FOR THE SULFATION AND ACTIVATION OF MINOXIDIL. MEDIATES THE METABOLIC ACTIVATION OF CARCINOGENIC N-HYDROXYARYLAMINES TO DNA BINDING PRODUCTS AND COULD SO PARTICIPATE AS MODULATING FACTOR OF CANCER RISK.
 CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + a phenol = adenosine 3',5'-bisphosphate + an aryl sulfate.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- TISSUE SPECIFICITY: LIVER, LUNG, ADRENAL, BRAIN, PLATELETS AND SKIN.
 CC -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U09031; AAA18613.1; -
 CC EMBL; L19955; AAA02935.1; -
 CC EMBL; L10819; AAA35562.1; -
 CC EMBL; U26309; AAA67895.1; -
 CC EMBL; L19999; AAA99892.1; -
 CC EMBL; X84654; CAA59147.1; -

DR EMBL; X78283; CAA55089.1; -;
 DR EMBL; U54701; AAC50480.1; -;
 DR EMBL; L5346; AAC60595.1; -;
 DR EMBL; U52852; AAC51816.1; -;
 DR EMBL; AJ007418; CAA07495.1; -;
 DR EMBL; BC000923; AAH00923.1; -;
 DR EMBL; U71086; AAB09597.1; -;
 DR HSP; P50224; ICMJ.
 DR Genew; HGNC:11453; SULT1A1.
 DR MIM; 171150; -;
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 KW Transferase; Steroid metabolism; Polymorphism.
 FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).
 FT VARIANT 37 37 R -> Q.
 FT VARIAT 213 213 R -> H (IN SULT1A1*2; FREQUENT
 POLYMORPHISM; HAS LOWER ACTIVITY).
 FT VARIAT 223 223 /FTID=VAR_007425.
 FT VARIAT 223 223 M -> V.
 FT MUTAGEN 70 70 /FTID=VAR_009303.
 C-S; INCREASED SENSITIVITY OF ENZYME
 ACTIVITY TO HEAT INACTIVATION.
 FT CONFLICT 90 90 P -> L (IN REF. 3).
 FT CONFLICT 146 146 A -> T (IN REF. 5).
 FT CONFLICT 181 181 E -> G (IN REF. 5).
 FT CONFLICT 243 243 V -> A (IN REF. 3).
 FT CONFLICT 282 282 E -> K (IN REF. 4 AND 8).
 FT CONFLICT 290 290 S -> T (IN REF. 4 AND 8).
 SQ SEQUENCE 295 AA; 34197 MW; 60E9D5BBA9159176 CRC64;
 Query Match 16.98; Score 54; DB 1; Length 295;
 Best Local Similarity 69.2%; Pred. No. 0.0056;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 220 RKGXGDKXKFT 232
 DB 257 RKGXGDKXKFT 269
 RESULT 4
 SUP2 HUMAN STANDARD; PRT; 295 AA.
 AC P50226; P78393;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phenol-sulfating phenol sulfotransferase 2 (EC 2.8.2.1) (P-PST)
 DE (S1A2).
 GN SULT1A2 OR STP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Zhu X., Veronese M.E., Iocco P., McManus M.E.;
 RN Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Kato R.;
 RN MEDLINE=96065417; PubMed=7581483;
 RA Ozawa S., Nagata K., Shimada M., Ueda M., Tsuruki T., Yamazoe Y.,
 RT "Primary structures and properties of two related forms of aryl
 sulfotransferases in human liver.";
 RL Pharmacogenetics 5:5135-5140(1995).
 [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94306556; PubMed=8033246;
 RA Yamazoe Y., Nagata K., Ozawa S., Kato R.;
 RT "Structural similarity and diversity of sulfotransferases.";
 Chem. Biol. Interact. 92:107-117(1994).
 [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96299636; PubMed=8661000;
 RA Her C., Rattiganis R., Weinshilboum R.M.;
 RT "Human phenol sulfotransferase STP2 gene: molecular cloning,
 structural characterization, and chromosomal localization.";
 Genomics 33:409-420(1996).
 [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97237042; PubMed=9119390;
 RA Gaedigk A., Beatty B.G., Grant D.M.;
 RT "Cloning, structural organization, and chromosomal mapping of the
 human phenol sulfotransferase STP2 gene.";
 Genomics 40:242-246(1997).
 [7]
 RP CHARACTERIZATION.
 RX MEDLINE=95196680; PubMed=7889867;
 RA Yamazoe Y., Ozawa S., Nagata K., Gong D.-W., Kato R.;
 RT "Characterization and expression of hepatic sulfotransferase involved
 in the metabolism of N-substituted aryl compounds.";
 Environ. Health Perspect. 102:99-103(1994).
 [8]
 RP VARIANT ASN-235.
 RX MEDLINE=20222641; PubMed=10762004;
 RA Engelke C.E., Meinel W., Boeing H., Glatt H.;
 RT "Association between functional genetic polymorphisms of human
 sulfotransferases 1A1 and 1A2.";
 Pharmacogenetics 10:163-169(2000).
 CC -1- FUNCTION: CATALYZES THE SULFATE CONJUGATION OF CATECHOLAMINES, THE
 PHENOLIC DRUGS AND NEUROTRANSMITTERS. IS ALSO RESPONSIBLE FOR THE
 SULFATION AND ACTIVATION OF MINOXIDIL. MEDIATES THE METABOLIC
 ACTIVATION OF CARCINOGENIC N-HYDROXYARYLAMINES TO DNA BINDING
 PRODUCTS AND COULD SO PARTICIPATE AS MODULATING FACTOR OF CANCER
 RISK.
 CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + a phenol =
 adenosine 3',5'-bisphosphate + an aryl sulfate.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC EMBL; U28170; AAB09659.1; -;
 CC EMBL; U28169; AAB09658.1; -;
 CC EMBL; X78282; CAA55088.1; -;
 CC EMBL; U34804; AAB09758.1; -;
 CC EMBL; U72202; AAB08970.1; -;
 CC EMBL; U72196; AAB08970.1; JOINED.
 CC EMBL; U72197; AAB08970.1; JOINED.
 CC EMBL; U72198; AAB08970.1; JOINED.
 CC EMBL; U72199; AAB08970.1; JOINED.
 CC EMBL; U72200; AAB08970.1; JOINED.
 CC EMBL; U72201; AAB08970.1; JOINED.
 CC EMBL; U76619; AAB18753.1; -;
 CC EMBL; U33886; AAC51149.1; -;
 CC HSP; P50224; ICMJ.
 CC Genew; HGNC:11454; SULT1A2.
 CC MIM; 601292; -;
 CC InterPro; IPR000863; Sulfotransferase.
 CC Pfam; PF00685; Sulfotransferase; 1.
 CC ProDom; PD001218; Sulfotransferase; 1.

KW Transferase; Steroid metabolism; Polymorphism.
 FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).
 FT VARIANT 7 T -> I.
 FT 7 /FTID=VAR_007426.
 FT 235 235 T -> N.
 FT 19 19 /FTID=VAR_007427.
 FT CONFLICT 19 P -> L (IN REF. 2).
 FT CONFLICT 290 290 S -> T (IN REF. 6).
 SQ SEQUENCE 295 AA; 34285 MW; FAD88E348B49BE5A CRC64;

Query Match 16.9%; Score 54; DB 1; Length 295;
 Best Local Similarity 69.2%; Pred. NO. 0.0056;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 220 RKGXXGDWCKXFT 232
 DB 257 RKGAGDWKTFT 269

RESULT 5

SUPM_HUMAN STANDARD; PRT; 295 AA.
 AC P50224;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Monamine-sulfating phenol sulfotransferase (EC 2.8.2.1)
 DE (Sulfotransferase, monamine-prefering) (M-PST) (Thermolabile phenol
 DE sulfotransferase) (PL-PST) (Placental estrogen sulfotransferase)
 DE (Catecholamine-sulfating phenol sulfotransferase) (HAST3).
 GN SULT1A3 OR STM.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93371391; PubMed=8363592;
 RA Zhu X., Veronese M.E., Bernard C.C., Sansom L.N., McManus M.E.;
 RT "Identification of two human brain aryl sulfotransferase cDNAs";
 RL Biochem. Biophys. Res. Commun. 195:120-127(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=94244843; PubMed=8187949;
 RA Bernier F., Lopez-Solache I., Labrie F., Luu-The V.;
 RT "Cloning and expression of cDNA encoding human placental estrogen
 RT sulfotransferase";
 RL Mol. Cell. Endocrinol. 99:R11-R15(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95100963; PubMed=7802665;
 RA Dooley T.P., Probst P., Munroe P.B., Mole S.E., Liu Z., Doggett N.A.;
 RT "Genomic organization and DNA sequence of the human catecholamine-
 RT sulfating phenol sulfotransferase gene (STM)";
 RL Biochem. Biophys. Res. Commun. 205:1325-1332(1994).
 RN [4]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 84-101.
 RC TISSUE=Liver;
 RX MEDLINE=94161723; PubMed=8117269;
 RA Wood T.C., Aksoy I.A., Aksoy S., Weinshilboum R.M.;
 RT "Human liver thermolabile phenol sulfotransferase: cDNA cloning,
 RT expression and characterization";
 RL Biochem. Biophys. Res. Commun. 198:1119-1127(1994).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95209696; PubMed=7695637;
 RA Aksoy I.A., Weinshilboum R.M.;
 RT "Human thermolabile phenol sulfotransferase gene (STM): molecular
 RT cloning and structural characterization";
 RL Biochem. Biophys. Res. Commun. 208:786-795(1995).

RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=95209704; PubMed=7695643;
 RA Jones A.L., Hagen M., Coughtrie M.W.H., Roberts R.C., Glatt H.;
 RT "Human platelet phenolsulfotransferases: cDNA cloning, stable
 RT expression in V79 cells and identification of a novel allelic variant
 RT of the phenol-sulfating form";
 RL Biochem. Biophys. Res. Commun. 208:855-862(1995).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leukocyte;
 RX MEDLINE=95050600; PubMed=7961757;
 RA Bernier F., Leblanc G., Labrie F., Luu-The V.;
 RT "Structure of human estrogen and aryl sulfotransferase gene. Two mRNA
 RT species issued from a single gene";
 RL J. Biol. Chem. 269:28200-28205(1994).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP SEQUENCE OF 139-198 FROM N.A.
 RC TISSUE=Lymphocytes;
 RX MEDLINE=95130098; PubMed=7829089;
 RA Aksoy I.A., Callen D.F., Apostolou S., Her C., Weinshilboum R.M.;
 RT "Thermolabile phenol sulfotransferase gene (STM): localization to
 RT human chromosome 16p11.2";
 RL Genomics 23:275-277(1994).
 RN [10]
 RP CHARACTERIZATION.
 RX MEDLINE=94379981; PubMed=8093002;
 RA Veronese M.E., Burgess W., Zhu X., McManus M.E.;
 RT "Functional characterization of two human sulphotransferase cDNAs
 RT that encode monamine- and phenol-sulphating forms of phenol
 RT sulphotransferase: substrate kinetics, thermal-stability and
 RT inhibitor-sensitivity studies";
 RL Biochem. J. 302:497-502(1994).
 RN [11]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RC TISSUE=Brain;
 RX MEDLINE=20013065; PubMed=10543947;
 RA Bidwell L.M., McManus M.E., Gaedigk A., Kakuta Y., Negishi M.,
 RA Pedersen L., Martin J.L.;
 RT "Crystal structure of human catecholamine sulfotransferase";
 RL J. Mol. Biol. 293:521-530(1999).
 CC -!- FUNCTION: CATALYZES THE SULFATE CONJUGATION OF PHENOLIC MONOAMINES
 CC (NEUROTRANSMITTERS SUCH AS DOPAMINE, NOREPINEPHRINE AND SEROTONIN)
 CC AND PHENOLIC AND CATECHOL DRUGS.
 CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + a phenol =
 CC adenosine 3',5'-bisphosphate + an aryl sulfate.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- TISSUE SPECIFICITY: LIVER, COLON, KIDNEY, LUNG, BRAIN, SPLEEN,
 CC SMALL INTESTINE, PLACENTA AND LEUKOCYTE.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L19956; AAA02943.1; -
 CC EMBL; L25275; AAA36523.1; -
 CC EMBL; U08032; AAAL7723.1; -
 CC EMBL; U20499; AAA64490.1; -
 CC EMBL; X84653; CAA59146.1; -
 CC EMBL; L34160; -; NOT_ANNOTATED_CDS.

```

DR EMBL; U37686; AAA86536.1; -.
DR EMBL; BC014471; AAH14471.1; -.
DR EMBL; U08099; AA82126.1; -.
DR PDB; 1CJM; 10-NOV-99.
DR Genew; HGNC:11455; SULT1A3.
DR MIM; 600641; -.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid metabolism; 3D-structure. (POTENTIAL).
FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).
SQ SEQUENCE 295 AA; 34196 MW; ECDC03DBE30D46 CRC64;

Query Match 16.9%; Score 54; DB 1; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.0056;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGGXGDWKKXXFT 232
DB 257 RKGWGDWKTFT 269

RESULT 6
SUPP_MACFA STANDARD; PRT; 295 AA.
AC P52846;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phenol-sulfating phenol sulfotransferase (EC 2.8.2.1) (P-PST).
GN STP.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Carcophthecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Ogura K., Satsukawa M., Okuda H., Watabe T.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE SULFATE CONJUGATION OF CATECHOLAMINES AND
CC PHENOLIC DRUGS AND NEUROTRANSMITTERS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + a phenol =
CC adenosine 3',5'-bisphosphate + an aryl sulfate.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D85514; BAA12822.1; -.
CC HSSP; P50224; 1CJM.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid metabolism.
FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).
SQ SEQUENCE 295 AA; 34238 MW; BD51639D1570A841 CRC64;

Query Match 16.9%; Score 54; DB 1; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.0056;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGGXGDWKKXXFT 232
DB 257 RKGWGDWKTFT 269

```

```

RESULT 7
SUH2_MOUSE STANDARD; PRT; 285 AA.
AC P50236;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alcohol sulfotransferase 2 (EC 2.8.2.2) (Hydroxysteroid
DE sulfotransferase) (ST).
GN STA2 OR STH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=94306561; PubMed=8033251;
RA Kong A.-N.T., Fei P.;
RL Molecular cloning of three sulfotransferase cDNAs from mouse liver.
RL Chem. Biol. Interact. 92:161-168(1994).
CC -1- FUNCTION: CATALYZES SULFATION OF HYDROXYSTEROIDS AND XENOBIOTICS
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + an alcohol =
CC adenosine 3',5'-bisphosphate + an alkyl sulfate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L27121; AAA40145.1; -.
CC HSSP; P50224; 1CJM.
CC MGD; MGI:107550; Sth2.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid metabolism.
FT BINDING 249 255 PAPS BINDING SITE (POTENTIAL).
SQ SEQUENCE 285 AA; 33329 MW; C49D138FE2B04308 CRC64;

Query Match 16.6%; Score 53; DB 1; Length 285;
Best Local Similarity 69.2%; Pred. No. 0.0084;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGGXGDWKKXXFT 232
DB 247 RKGTTGDWKNHFT 259

RESULT 8
SUH2_RAT STANDARD; PRT; 285 AA.
AC P07631; O09038;
DT 01-APR-1988 (Rel. 07, Created)
DT 30-WAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable alcohol sulfotransferase (EC 2.8.2.2) (Hydroxysteroid
DE sulfotransferase) (ST) (Senescence marker protein 2) (SMP-2)
DE sulfotransferase) (DST).
DE (Androgen-repressible liver protein) (dehydroepiandrosterone
DE sulfotransferase) (DST).
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]

```


CC -I- INDUCTION: INDUCED BY ESTROGENS AND SUPPRESSED BY ANDROGENS.
CC EXPRESSION IS UNDER THE INFLUENCE OF PITUITARY GROWTH HORMONE AND
CC THYROID HORMONE.
CC
CC -I- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR ENBL; M31363; AAA41356.1; -
DR ENBL; D14988; BAA03633.1; -
DR ENBL; D14987; BAA03632.1; -
DR PIR; A33569; A33569.
DR HSSP; P50224; ICJM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid metabolism.
FT INIT_MET 0
FT BINDING 247 253 PAPS BINDING SITE (POTENTIAL).
FT CONFLICT 32 32 E -> D (IN REF. 1).
FT CONFLICT 118 118 I -> V (IN REF. 1).
FT CONFLICT 138 138 V -> A (IN REF. 1).
FT CONFLICT 172 172 R -> Q (IN REF. 1).
FT CONFLICT 241 241 I -> T (IN REF. 1).
FT CONFLICT 249 249 I -> V (IN REF. 1).
SQ SEQUENCE 283 AA; 33059 MW; 619D72398D6CE3DC CRC64;

Query Match 16.2%; Score 52; DB 1; Length 283;
Best Local Similarity 69.2%; Pred. No. 0.013;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGGXGDWKKXFT 232
Db 245 RKGTIGDWKNHFT 257

RESULT 11
ID SUHA HUMAN STANDARD; PRT; 284 AA.
AC Q06520;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alcohol sulfotransferase (EC 2.8.2.2) (Hydroxysteroid
DE Sulfotransferase) (HST) (Dehydroepiandrosterone sulfotransferase)
DE (DHEA-ST) (ST2) (ST2A3).
GN SULT2A1 OR STD OR HST.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 60-64; 104-119 AND 273-284.
RC TISSUE=Liver;
RX MEDLINE=93143674; PubMed=7678732;
RA Cower K.A., Falany J.L., Falany C.N.;
RT "Cloning and expression of human liver dehydroepiandrosterone
RT sulfotransferase."
RL Biochem. J. 289:233-240(1993).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 80-107 AND 176-198.
RC TISSUE=Liver;
RX MEDLINE=92269778; PubMed=1588921;
RA Otterness D.M., Wieben E.D., Wood T.C., Madden B.J.,
RA McCormick D.J., Weinshilboum R.M.;
RT "Human liver dehydroepiandrosterone sulfotransferase: molecular
RT cloning and expression of cDNA."
RL Mol. Pharmacol. 41:865-872(1992).

RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RX MEDLINE=96034512; PubMed=7589785;
RA Forbes K.J., Hagen M., Coughtrie M.W.H., Glatt H.R., Hume R.;
RT "Human fetal adrenal hydroxysteroid sulphotransferase: cDNA cloning,
RT stable expression in V79 cells and functional characterisation of the
RT expressed enzyme."
RL Mol. Cell. Endocrinol. 112:53-60(1995).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=95322029; PubMed=7598806;
RA Luu-The V., Dufort I., Paquet N., Reimnitz G., Labrie F.;
RT "Structural characterization and expression of the human
RT dehydroepiandrosterone sulfotransferase gene."
RL DNA Cell Biol. 14:511-518(1995).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=95225980; PubMed=7710689;
RA Otterness D.M., Her C., Aksoy S., Kimura S., Wieben E.D.,
RA Weinshilboum R.M.;
RT "Human dehydroepiandrosterone sulfotransferase gene: molecular
RT cloning and structural characterization."
RL DNA Cell Biol. 14:331-341(1995).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=92392364; PubMed=1520333;
RA Kong A.-N.T., Yang L., Ma M., Rao D., Bjornsson T.D.;
RT "Molecular cloning of the alcohol/hydroxysteroid form (hSta) of
RT sulfotransferase from human liver."
RL Biochem. Biophys. Res. Commun. 187:448-454(1992).
RN [7]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: CATALYZES THE SULFATION OF STEROIDS AND BILE ACIDS IN
CC THE LIVER AND ADRENAL GLANDS.
CC -I- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + an alcohol =
CC adenosine 3',5'-bisphosphate + an alkyl sulfate.
CC -I- SUBUNIT: HOMODIMER.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- TISSUE SPECIFICITY: LIVER, ADRENAL AND AT LOWER LEVEL IN THE
CC KIDNEY. IS PRESENT IN HUMAN FETUS IN HIGHER LEVEL IN THE ADRENAL
CC THAN THE LIVER AND THE KIDNEY.
CC -I- PTM: THE N-TERMINUS IS BLOCKED.
CC -I- MISCELLANEOUS: ESTROGENS PRESENT IN MATERNAL CIRCULATION IS
CC PREDOMINANTLY DERIVED FROM FETAL DEHYDROPIANDOSTERONE SULFATE
CC WHICH IS HYDROLYZED AND METABOLIZED TO ESTROGENS IN PLACENTA.
CC -I- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR ENBL; L20000; AAA35758.1; -
DR ENBL; X70222; CAA49755.1; -
DR ENBL; U08024; AAA17749.1; -
DR ENBL; U08025; AAA17750.1; -
DR ENBL; X84816; CAA59274.1; -
DR ENBL; L36196; AAA75491.1; -
DR ENBL; L36191; AAA75491.1; JOINED.
DR ENBL; L36192; AAA75491.1; JOINED.
DR ENBL; L36193; AAA75491.1; JOINED.
DR ENBL; L36194; AAA75491.1; JOINED.
DR ENBL; L36195; AAA75491.1; JOINED.
DR ENBL; U13061; AAC51353.1; -
DR ENBL; U13056; AAC51353.1; JOINED.

```
DR EMBL; U13057; AAC51353.1; JOINED.
DR EMBL; U13058; AAC51353.1; JOINED.
DR EMBL; U13059; AAC51353.1; JOINED.
DR EMBL; U13060; AAC51353.1; JOINED.
DR EMBL; S43859; AAB23169.2; -.
DR EMBL; BC020755; AAH20755.1; -.
DR HSP; P50224; 1CJM.
DR Genew; HGNC:11458; SULT2A1.
DR MIM; 125263; -.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid metabolism.
FT BINDING 248 254 PAPS BINDING SITE (POTENTIAL).
FT CONFLICT 62 62 A -> P (IN REF. 1; AA SEQUENCE).
FT CONFLICT 89 89 T -> S (IN REF. 1).
FT CONFLICT 118 118 L -> D (IN REF. 1; AA SEQUENCE).
FT CONFLICT 158 158 L -> V (IN REF. 6).
SQ SEQUENCE 284 AA; 33648 MW; 3C89C7597833EBAA1 CRC64;

Query Match 16.2%; Score 52; DB 1; Length 284;
Best Local Similarity 69.2%; Pred. No. 0.013;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGGXGDMKXXFT 232
DB 246 RKGVGDWKNHFT 258

RESULT 13
SUHA_CAVPO
ID SUHA_CAVPO STANDARD; PRT; 286 AA.
AC P50234;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 3-alpha-hydroxysteroid sulfotransferase (EC 2.8.2.2) (Alcohol
DE sulfotransferase) (HStr1).
GN STD1 OR STD.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Caviidae; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH 2; TISSUE=Adrenal gland;
RX MEDLINE=94253178; PubMed=8195238;
RA Lee Y.C., Park C.-S., Strodt C.A.;
RT "Molecular cloning of a chiral-specific 3 alpha-hydroxysteroid
RT sulfotransferase.";
RL J. Biol. Chem. 269:15838-15845 (1994).
CC -!- FUNCTION: CATALYSES THE SULFATION OF 3-ALPHA-HYDROXYL GROUPS OF
CC NEUTRAL STEROIDS.
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + an alcohol =
CC adenosine 3',5'-bisphosphate + an alkyl sulfate.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- TISSUE SPECIFICITY: ADRENAL GLANDS AND LIVER.
CC -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U06871; AAA19588.1; -.
DR HSP; P50224; 1CJM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid metabolism.
FT INIT_MET 0 0 BY SIMILARITY.
FT BINDING 248 254 PAPS BINDING SITE (POTENTIAL).
SQ SEQUENCE 286 AA; 33508 MW; 2608A8D7F326869A CRC64;

Query Match 16.2%; Score 52; DB 1; Length 286;
Best Local Similarity 69.2%; Pred. No. 0.013;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGGXGDMKXXFT 232
DB 246 RKGVGDWKNHFT 258

RESULT 14
SUOE_HUMAN
ID SUOE_HUMAN STANDARD; PRT; 285 AA.
AC P52843;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alcohol sulfotransferase 1 (EC 2.8.2.2) (Hydroxysteroid
DE sulfotransferase) (ST).
GN SULT2A1 OR STAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93248107; PubMed=8493851;
RA Kong A.-N.T., Tao D., Ma M., Yang L.;
RT "Molecular cloning of the alcohol/hydroxysteroid form (mStal) of
RT sulfotransferase from mouse liver.";
RL Pharm. Res. 10:627-630 (1993).
CC -!- FUNCTION: CATALYSES SULFATION OF HYDROXYSTEROIDS AND XENOBIOTICS
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + an alcohol =
CC adenosine 3',5'-bisphosphate + an alkyl sulfate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L02335; -. NOT ANNOTATED_CDS.
DR HSP; P50224; 1CJM.
DR MGD; MGI:98430; Sult2a1.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
```

ID SUOE HUMAN STANDARD; PRT; 294 AA.
 AC P49828;
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Estrogen sulfotransferase (EC 2.8.2.4) (Sulfotransferase, estrogen-
 DE preferring) (EST-1).
 GN SUL1E1 OR STE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94242031; PubMed=8185618;
 RA Aksoy I.A., Wood T.C., Weinshilboum R.;
 RT "Human liver estrogen sulfotransferase: identification by cDNA
 RT cloning and expression.";
 RL Biochem. Biophys. Res. Commun. 200:1621-1629 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=96079087; PubMed=8530066;
 RA Her C., Aksoy I.A., Kimura S., Brandriff B.F., Wasmuth J.J.,
 RA Weinshilboum R.M.;
 RT "Human estrogen sulfotransferase gene (STE): cloning, structure, and
 RT chromosomal localization.";
 RL Genomics 29:16-23 (1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95298714; PubMed=7779757;
 RA Falany C.N., Krasnykh V., Falany J.L.;
 RT "Bacterial expression and characterization of a cDNA for human liver
 RT estrogen sulfotransferase.";
 RL J. Steroid Biochem. Mol. Biol. 52:529-539 (1995).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Rubin G.L.;
 RN Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 182-294 FROM N.A.
 RC TISSUE=Liver, and Spleen;
 RA Her C., Szumlanski C., Aksoy I., Weinshilboum R.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY CONTROL THE LEVEL OF THE ESTROGEN RECEPTOR BY
 CC SULFYLATING FREE ESTRADIOL. MAXIMALLY SULFATES BETA-ESTRADIOL
 CC AND ESTRONE AT CONCENTRATIONS OF 20 NM. ALSO SULFATES
 CC DEHYDROEPIANDROSTERONE, PREGNENOLONE, ETHINYLESTRADIOL, EQUALININ,
 CC DIETHYLSTILBESTROL AND 1-NAPHTHOL, AT SIGNIFICANTLY HIGHER
 CC CONCENTRATIONS; HOWEVER, CORTISOL, TESTOSTERONE AND DOPAMINE ARE
 CC NOT SULFATED.
 CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone =
 CC adenosine 3',5'-bisphosphate + estrone 3-sulfate.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: LIVER, INTESTINE AND AT LOWER LEVEL IN THE
 CC KIDNEY.
 CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.iesb-sib.ch/announce/>
 CC or send an email to license@iesb-sib.ch).
 CC
 CC EMBL; U08098; AAA82125.1; -
 CC EMBL; U20521; AAC50286.1; -
 CC EMBL; U20515; AAC50286.1; JOINED.
 CC EMBL; U20516; AAC50286.1; JOINED.

DR EMBL; U20517; AAC50286.1; JOINED.
 DR EMBL; U20518; AAC50286.1; JOINED.
 DR EMBL; U20519; AAC50286.1; JOINED.
 DR EMBL; U20520; AAC50286.1; JOINED.
 DR EMBL; S77383; AAB34601.1; -
 DR EMBL; Y11195; CAA72079.1; -
 DR EMBL; U55764; AAB51658.1; -
 DR HSP; P49891; 1AQU.
 DR Genew; HGNC:11377; STE.
 DR MIM; 600043; -
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 KW Transferase; Steroid-binding.
 FT BINDING 258 264
 FT PAPS BINDING SITE (POTENTIAL).
 SQ SEQUENCE 294 AA; 35126 MW; 9EC8923D20757D57 CRC64;

 Query Match 16.2%; Score 52; DB 1; Length 294;
 Best Local Similarity 69.2%; Pred. No. 0.013;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

 QY 220 RKGXXGDKKXET 232
 DB 256 RKGITGDKWKNHFT 268

 RESULT 15:
 SICL_HUMAN STANDARD; PRT; 296 AA.
 ID SICL_HUMAN STANDARD; PRT; 296 AA.
 AC 000338;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sulfotransferase 1C1 (EC 2.8.2.-) (SULT1C1) (SULT1C2) (humsULTC2).
 GN SULT1C1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal liver, and Fetal spleen;
 RX MEDLINE=97312707; PubMed=9189148;
 RA Her C., Kaur G.P., Athwal R.S., Weinshilboum R.M.;
 RT "Human sulfotransferase SULT1C1: cDNA cloning, tissue-specific
 RT expression, and chromosomal localization.";
 RL Genomics 41:467-470 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal liver;
 RX MEDLINE=98297811; PubMed=9635888;
 RA Yoshinari K., Nagata K., Shimada M., Yamazoe Y.;
 RT "Molecular characterization of SULT1C1-related human sulfotransferase.";
 RL Carcinogenesis 19:951-953 (1998).
 RN [3]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Fetal lung;
 RX MEDLINE=99069375; PubMed=9852044;
 RA Sakatibara Y., Yanagisawa K., Katafuchi J., Ringer D.P., Takami Y.,
 RA Nakayama T., Suiko M., Liu M.-C.;
 RT "Molecular cloning, expression, and characterization of novel human
 RT SULT1C sulfotransferases that catalyze the sulfonation of
 RT N-hydroxy-2-acetylaminofluorene.";
 RL J. Biol. Chem. 273:33929-33935 (1998).
 RN [4]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Stomach;
 RX MEDLINE=99410886; PubMed=10481272;
 RA Hehonah N., Zhu X., Brix L., Bolton-Grob R., Barnett A., Windmill K.,
 RA McManus M.;
 RT "Molecular cloning, expression, localisation and functional
 RT characterisation of a rabbit SULT1C2 sulfotransferase.";
 RL Int. J. Biochem. Cell Biol. 31:869-882 (1999).

Job time : 24 secs

[5]
 RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=20247255; PubMed=10783263;
 RA Fremuth R.R., Rafogiannis R.B., Wood T.C., Moon E., Kim U.-J., Xu J.,
 RT Siciliano M.J., Weinshilboum R.M.;
 RA "Human sulfotransferases SULT1C1 and SULT1C2: cDNA characterization,
 RT gene cloning, and chromosomal localization.";
 RL Genomics 65:157-165(2000).
 [6]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Kidney;
 RC Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE SULFATE CONJUGATION OF MANY DRUGS,
 CC XENOBIOTIC COMPOUNDS, HORMONES, AND NEUROTRANSMITTERS. MAY BE
 CC INVOLVED IN THE ACTIVATION OF CARCINOGENIC HYDROXYLAMINES. SHOWS
 CC ACTIVITY TOWARDS P-NITROPHENOL AND N-HYDROXY-2-ACETYLAMINO-
 CC FLUORENE (N-OH-2AAF).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A SHORT FORM (SHOWN HERE) AND A
 CC LONG FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: FOUND IN ADULT STOMACH, KIDNEY AND THYROID
 CC GLAND, AND IN FETAL KIDNEY AND LIVER.
 CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U66036; AAC51285.1; -;
 DR EMBL; AB008164; BAA28346.1; -;
 DR EMBL; AF026303; AAC00409.1; -;
 DR EMBL; AF186251; AAF72799.1; -;
 DR EMBL; AF186252; AAF72800.1; -;
 DR EMBL; AF186253; AAF72801.1; -;
 DR EMBL; AF186254; AAF72802.1; -;
 DR EMBL; AF186255; AAF72803.1; -;
 DR EMBL; AF186256; AAF72804.1; -;
 DR EMBL; AF186262; AAF72805.1; -;
 DR EMBL; AF186258; AAF72805.1; JOINED.
 DR EMBL; AF186260; AAF72805.1; JOINED.
 DR EMBL; AF186261; AAF72805.1; JOINED.
 DR EMBL; AF186262; AAF72806.1; -;
 DR EMBL; AF186258; AAF72806.1; JOINED.
 DR EMBL; AF186259; AAF72806.1; JOINED.
 DR EMBL; AF186260; AAF72806.1; JOINED.
 DR EMBL; AF186261; AAF72806.1; JOINED.
 DR EMBL; BC005353; AAH05353.1; -;
 DR HSSP; P50224; 1CJM.
 DR Genew; HGNC:11456; SULT1C1.
 DR MIM; 602385; -;
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransferase; 1.
 DR Prodom; PD001218; Sulfotransferase; 1.
 KW Transferase; Alternative splicing.
 FT VARSPPLIC 93 113
 FT SGNPPASTSQSAKITD (IN LONG ISOFORM).
 SQ SEQUENCE 296 AA; 34880 MW; 3DC01C8A8ED61EFD CRC64;

Query Match 16.2%; Score 52; DB 1; Length 296;
 Best Local Similarity 69.2%; Pred. No. 0.014;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 220 RKGGXGDXKXFT 232
 DB 258 RKGTVGDWKNHFT 270

Search completed: June 5, 2003, 12:29:54

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:25:03 ; Search time 80 Seconds

(without alignments)
597.537 Million cell updates/sec

Title: US-09-854-122-1

Perfect score: 320

Sequence: 1 TYPKSGTXXXXXXX.....XXXXXXXXXXGDXGDXGDXFT 232

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:*

- 1: sp_archea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	16.9	42	4 Q9UMT9	Q9umt9 homo sapien
2	54	16.9	295	4 Q95603	Q95603 homo sapien
3	54	16.9	295	6 Q29476	Q29476 canis famill
4	53	16.6	268	11 Q9DCK7	Q9dck7 mus musculus
5	53	16.6	296	4 Q43704	Q43704 homo sapien
6	53	16.6	296	4 Q15497	O15497 homo sapien
7	53	16.6	296	4 Q96F11	Q96f11 homo sapien
8	53	16.6	298	11 Q9R1S5	Q9r1s5 mus musculus
9	52	16.2	249	16 Q9K710	Q9k710 bacillus ha
10	52	16.2	286	6 Q62648	O62648 oryctolagus
11	52	16.2	295	6 Q95MP8	Q95mfs sus scrofa
12	52	16.2	295	6 Q95JD6	Q95jd6 canis famill
13	52	16.2	295	11 Q35401	Q35401 mus musculus
14	52	16.2	296	6 Q95JD5	Q95jd5 canis famill
15	52	16.2	296	11 Q9D939	Q9d939 mus musculus
16	52	16.2	296	11 Q8R210	Q8r210 mus musculus

17	52	16.2	307	13 Q90WR6	Q90wr6 gallus gall
18	52	16.2	308	11 Q9Z1G0	Q9z1g0 rattus norv
19	52	16.2	309	11 Q9R2C2	Q9r2c2 mus musculu
20	52	16.2	338	11 Q35400	O35400 mus musculu
21	52	16.2	338	11 Q91V03	Q91v03 mus musculu
22	51	15.9	292	6 Q9XT99	Q9xt99 oryctolagus
23	51	15.9	295	6 Q95KM3	Q95km3 oryctolagus
24	51	15.9	295	6 Q95JCS	Q95jcs ornithorhyn
25	51	15.9	296	6 Q95JCS	Q95jcs trichosurus
26	51	15.9	299	11 Q9QMG7	Q9qmg7 mus musculu
27	51	15.9	299	11 Q9Z2T0	Q9z2t0 mus musculu
28	51	15.9	1313	5 Q8T1M7	Q8t1m7 dictyosteli
29	50	15.6	73	4 Q9UK34	Q9uk34 homo sapien
30	50	15.6	193	11 Q91XG4	Q91xg4 mus musculu
31	50	15.6	263	11 Q91W19	Q91w19 mus musculu
32	50	15.6	295	11 Q9D566	Q9d566 mus musculu
33	50	15.6	304	11 Q70262	O70262 mus musculu
34	50	15.6	350	4 Q00204	O00204 homo sapien
35	50	15.6	350	4 Q75814	Q75814 homo sapien
36	50	15.6	365	4 Q00205	O00205 homo sapien
37	49	15.3	67	11 Q63551	Q63551 rattus norv
38	49	15.3	67	11 Q04169	Q04169 rattus norv
39	49	15.3	293	11 Q35403	O35403 mus musculu
40	49	15.3	301	6 Q46640	O46640 oryctolagus
41	46	14.4	291	11 Q35402	Q35402 mus musculu
42	46	14.4	291	11 Q91X36	Q91x36 mus musculu
43	46	14.4	295	11 Q9QWS0	Q9qws0 rattus norv
44	46	14.4	295	11 Q99ND5	Q99nds rattus norv
45	46	14.4	331	10 Q9FX56	Q9fx56 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q9UMT9	PRELIMINARY;	PRT;	42 AA.
AC	Q9UMT9;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DE	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Phenol sulfotransferase (Fragment).			
GN	STP1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96299636; PubMed=8661000;			
RA	Her C., Raftogianis R., Weinshilboum R.M.;			
RT	"Human phenol sulfotransferase STP2 gene: molecular cloning, structural characterization, and chromosomal localization.";			
RL	Genomics 33:409-420(1996).			
DR	EMBL; U37025; AAB09759.1; -.			
DR	HSSP; P50224; 1CJM.			
DR	InterPro: IPR000863; Sulfotransferase.			
DR	Pfam; PF00685; Sulfotransferase; 1.			
DR	ProDom; PD001218; Sulfotransferase; 1.			
FT	NON TER			
FT	NON TER			
SQ	SEQUENCE	42 AA;	4794 MW;	87240072A7D3178D CRC64;

Query Match 16.9%; Score 54; DB 4; Length 42;

Best Local Similarity 69.2%; Pred. No. 0.0033; 4; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDXGDXFT 232

Db 4 RKGAGDWTFT 16

RESULT 2

```

O95603
ID O95603 PRELIMINARY; PRT; 295 AA.
AC O95603;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Aryl sulfotransferase.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Gaedigk A., Grant D.M.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U34199; AAC99987.1; -.
DR HS9P; P50224; 1CJM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 295 AA; 34219 MW; BC3411B083E2F7CB CRC64;

Query Match 16.9%; Score 54; DB 4; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.029; 4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Qy 220 RKGXGDKWXXFT 232
Db 257 RKGAGDKWTKTFT 269

RESULT 3
O29476 PRELIMINARY; PRT; 295 AA.
AC O29476;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Phenol sulfotransferase (EC 2.8.2.1) (Aryl sulfotransferase)
DE (Sulfokinase)
DE PST OR SULT1A1.
GN Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BEAGLE; TISSUE=LIVER;
RA Satsukawa M., Ogura K., Nakamura T., Watabe T.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 222-267 FROM N.A.
RA Liu P.-C., Shibuya H., Nonnen D., Katz M.L., Johnson G.S.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Tsai C., Morgenstern R., Swedmark S.;
RL Molecular cloning, expression, and characterization of canine
RT SULT1A1.
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLISULFATE + A PHENOL =
CC ADENOSINE 3',5'-BISPHOSPHATE + AN ARYL SULFATE.
DR EMBL; D29807; BAA06190.1; -.
DR EMBL; AF034534; AAB86976.1; -.
DR EMBL; AY069322; AAL57171.1; -.
DR HS9P; P50224; 1CJM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.

SQ SEQUENCE 295 AA; 34219 MW; BC3411B083E2F7CB CRC64;

Query Match 16.9%; Score 53; DB 11; Length 268;
Best Local Similarity 69.2%; Pred. No. 0.041; 4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Qy 220 RKGXGDKWXXFT 232
Db 230 RKGITGDKWTKTFT 242

RESULT 5
O43704 PRELIMINARY; PRT; 296 AA.
ID O43704;
AC O43704;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SULT1B2.
OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=98104061; PubMed=9443824;
 RA Fujita K., Nagata K., Ozawa S., Sasano H., Yamazoe Y.;
 RT "Molecular cloning and characterization of rat ST1B1 and human ST1B2
 RT cDNAs, encoding thyroid hormone sulfotransferases.";
 RL J. Biochem. 122:1052-1061(1997).
 DR EMBL; D89479; BAA24547.1; -.
 DR HSSP; P50224; 1CJM.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 SQ SEQUENCE 296 AA; 34827 MW; 9486076973859A8C CRC64;
 Query Match 16.6%; Score 53; DB 4; Length 296;
 Best Local Similarity 69.2%; Pred. No. 0.046;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 220 RKGXXGDWKKXFT 232
 DB 258 RKGTDGDKNYFT 270
 RESULT 6
 OI5497 PRELIMINARY; PRT; 296 AA.
 AC OI5497;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Thyroid hormone sulfotransferase.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Wang J., Falany J.L., Falany C.N.;
 RT "Expression and Characterization of a Novel Thyroid Hormone-Sulfating
 RT Form of Cytosolic Sulfotransferase From Human Liver.";
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U95726; AAB65154.1; -.
 DR HSSP; P50224; 1CJM.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 KW Transferase.
 SQ SEQUENCE 296 AA; 34899 MW; AFB61B21DBD782C CRC64;
 Query Match 16.6%; Score 53; DB 4; Length 296;
 Best Local Similarity 69.2%; Pred. No. 0.046;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 220 RKGXXGDWKKXFT 232
 DB 258 RKGTDGDKNYFT 270
 RESULT 7
 OI5497 PRELIMINARY; PRT; 296 AA.
 AC OI5497;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Unknown (protein for MGC:13356).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BONE MARROW;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC010895; AAH10895.1; -.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 SQ SEQUENCE 296 AA; 34927 MW; 8FE1EF99700BA352 CRC64;
 Query Match 16.6%; Score 53; DB 4; Length 296;
 Best Local Similarity 69.2%; Pred. No. 0.046;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 220 RKGXXGDWKKXFT 232
 DB 258 RKGTDGDKNYFT 270
 RESULT 8
 OI5497 PRELIMINARY; PRT; 298 AA.
 AC OI5497;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Arylsulfotransferase ST1A4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129SV;
 RA Honna W., Nagata K., Yamazoe Y.;
 RT "Mouse arylsulfotransferase ST1A4 gene.";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB029487; BAA82321.1; -.
 DR HSSP; P50224; 1CJM.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 KW Transferase.
 SQ SEQUENCE 298 AA; 34717 MW; E5100E784263668 CRC64;
 Query Match 16.6%; Score 53; DB 11; Length 298;
 Best Local Similarity 69.2%; Pred. No. 0.046;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 220 RKGXXGDWKKXFT 232
 DB 260 RKGTDGDKNYFT 272
 RESULT 9
 OI5497 PRELIMINARY; PRT; 249 AA.
 AC OI5497;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein BH3381.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;

RA MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001518; BAB07100.1; -
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 249 AA; 29291 MW; BA0902DDDFC7B9E CRC64;

Query Match 16.2%; Score 52; DB 16; Length 249;
Best Local Similarity 69.2%; Pred. No. 0.059; 4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXXGDWKKXFT 232
Db 211 RKGSGDWKNYFT 223
||| ||||| ||
||| ||||| ||

RESULT 10
O62648 PRELIMINARY; PRT; 286 AA.
AC O62648;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 07, Last annotation update)
DE AST-RB2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=LIVER;
RX MEDLINE=98207066; PubMed=9538269;
RA Yoshinari K., Nagata K., Shiraga T., Iwasaki K., Hata T., Yamazoe Y.;
RT "Molecular cloning, expression, and enzymatic characterization of
RT rabbit hydroxysteroid sulfotransferase AST-RB2 (St2A8).";
RL J. Biochem. 123:740-746(1998).
DR EMBL; AB006053; BAA25387.1; -
DR HSP; P50224; 1CJM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1
DR ProDom; PD001218; Sulfotransferase; 1.
SQ SEQUENCE 286 AA; 33582 MW; FCHD76DB270E7751 CRC64;

Query Match 16.2%; Score 52; DB 6; Length 286;
Best Local Similarity 69.2%; Pred. No. 0.069; 4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXXGDWKKXFT 232
Db 249 RKGSGDWKNYFT 261
||| ||||| ||
||| ||||| ||

RESULT 11
Q95MF8 PRELIMINARY; PRT; 295 AA.
AC Q95MF8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Estrogen sulfotransferase.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.

RA Kim J.G., Vallet J.L., Rohrer G.A., Christenson R.K.;
RT "Characterization of porcine uterine estrogen sulfotransferase.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF389855; AAK72967.1; -
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 295 AA; 34940 MW; 7677C28A91B38167 CRC64;

Query Match 16.2%; Score 52; DB 6; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.071; 4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXXGDWKKXFT 232
Db 257 RKGAGDWKNYFT 269
||| ||||| ||
||| ||||| ||

RESULT 12
O95JD6 PRELIMINARY; PRT; 295 AA.
AC Q95JD6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sulfotransferase (Fragment).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=21285733; PubMed=11389699;
RA Tsoi C., Falany C.N., Morgenstern R., Swedmark S.;
RT "Identification of a new subfamily of sulphotransferases: cloning and
RT characterization of canine SUL1D1.";
RL Biochem. J. 356:891-897(2001).
DR EMBL; AY004331; RAF86582.1; -
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
FT NON TER 295
SQ SEQUENCE 295 AA; 35207 MW; 62A4C580E8CAF2B8 CRC64;

Query Match 16.2%; Score 52; DB 6; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.071; 4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 220 RKGXXGDWKKXFT 232
Db 257 RKGSGDWKNYFT 269
||| ||||| ||
||| ||||| ||

RESULT 13
O35401 PRELIMINARY; PRT; 295 AA.
AC O35401;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Amine N-sulfotransferase.
GN SULTN OR SULTN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98321187; PubMed=9647753;
RA Sakakibara Y., Yanagisawa K., Takami Y., Nakayama T., Suiko M.,

```

RA Liu M.-C.;
RT "Molecular cloning, expression, and functional characterization of
RT novel mouse sulfotransferases.";
RL Biochem. Biophys. Res. Commun. 247:681-686(1998).
DR HMBL; AF026073; AAC69919.1; -.
DR HSSP; P50224; 1CJM.
DR MGD; MGI:1926341; Sultn.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 295 AA; 35099 MW; 17FE3E02AD6E1269 CRC64;

Query Match 16.2%; Score 52; DB 11; Length 295;
Best Local Similarity 69.2%; Pred. No. 0.071;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDKWKKXFT 232
DB 257 RKGISGDKWKNQFT 269

RESULT 14
Q95JD5 PRELIMINARY; PRT; 296 AA.
AC Q95JD5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sulfotransferase ST1B2 (Fragment).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=21261348; PubMed=11368519;
RA Teoi C., Falany C.N., Morgenstern R., Swedmark S.;
RT "Molecular cloning, expression, and characterization of a canine
RT Sulfotransferase that is a human ST1B2 ortholog.";
RL Arch. Biochem. Biophys. 390:87-92(2001).
DR EMBL; AY004332; AAF86583.1; -.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
FT NON TER
SQ SEQUENCE 296 AA; 34869 MW; 3D54CF003BF09365 CRC64;

Query Match 16.2%; Score 52; DB 6; Length 296;
Best Local Similarity 69.2%; Pred. No. 0.072;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDKWKKXFT 232
DB 258 RKGIAGDKWKNYFT 270

RESULT 15
Q9D939 PRELIMINARY; PRT; 296 AA.
AC Q9D939;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1810008N17Rik protein (Sulfotransferase).
GN SULT1C1 OR 1810008N17RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Sugimura K.;
RT "cDNA cloning of a novel mouse kidney sulfotransferase associated with
RT DPT-induced polycystic kidney.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK007384; BAB25002.1; -.
DR EMBL; AY005469; AAG00823.1; -.
DR HSSP; P50224; 1CJM.
DR MGD; MGI:1916333; Sult1c1.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 296 AA; 34953 MW; 50B8648F2A2BEA57 CRC64;

Query Match 16.2%; Score 52; DB 11; Length 296;
Best Local Similarity 69.2%; Pred. No. 0.072;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 220 RKGXGDKWKKXFT 232
DB 258 RKGTGDKWKNHFT 270

Search completed: June 5, 2003, 12:31:22
Job time : 82 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 17:17:46 ; Search time 3133 Seconds
(without alignments)
11072.628 Million cell updates/sec

Title: US-09-854-122-15
Perfect score: 1192
Sequence: 1 acggggggaataactggaat.....gttaaaaaaaaaaaaaa 1192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pin.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1192	100.0	1192	6	AX338910	Sequence
C 2	145.4	12.2	88010	8	AT114D3	AX338910 Sequence
C 3	145.4	12.2	89214	8	ATF14D17	AL138649 Arabidops
4	131.4	11.0	98198	2	AC119410	AL333992 Arabidops
5	130.2	10.9	102183	8	AC027134	AL119410 Medicago
6	130.2	10.9	103157	8	AC011810	AC027134 Arabidops
7	128.6	10.6	1212	8	AY085411	AC011810 Arabidops
8	127.2	10.7	996	8	AF325065	AY085411 Arabidops
9	127.2	10.7	1154	8	AY087921	AF325065 Arabidops
10	127.2	10.7	1216	8	AY092961	AY087921 Arabidops
C 11	127.2	10.7	87543	8	AC006836	AY092961 Arabidops
C 12	117.2	9.8	90341	8	AC006322	AC006836 Arabidops
C 13	117.2	9.8	118196	8	AC005824	AC006322 Arabidops
14	114	9.6	1179	8	AF375458	AC005824 Arabidops
C 15	111.8	9.4	121720	8	AC021044	AF375458 Arabidops
C 16	110.6	9.3	981	6	AX412507	AC021044 Arabidops
17	110.6	9.3	981	6	AX412507	AX412507 Sequence
18	110.6	9.3	981	8	AV113050	AX412507 Sequence
19	108.6	9.1	87286	8	AB026851	AY113050 Arabidops
C 20	108.6	9.1	89504	8	ATT25K17	AB026851 Arabidops
C 21	108.6	9.1	196286	8	ATCHRIV64	AB025K17 Arabidops
C 22	108.6	9.1	198320	8	ATCHRIV65	AL049171 Arabidops
23	107.4	9.0	1173	8	ATTS4391	AL161564 Arabidops
C 24	103.6	8.7	112732	2	AC123976	AL161564 Arabidops
C 25	103.2	8.7	1281	8	AY088081	Z46823 A. thaliana
C 26	103.2	8.7	108387	8	AC016662	AC123976 Medicago
27	102.4	8.6	1273	8	AY084999	AY088081 Arabidops
C 28	102.4	8.6	98198	2	AC119410	AC016662 Arabidops
C 29	100.8	8.5	1077	8	AY080458	AY084999 Arabidops
C 30	100.8	8.5	1347	8	AY099809	AC119410 Medicago
C 31	100.8	8.5	86380	8	AB010897	AX080458 Sequence
32	99.2	8.3	1270	8	AY054219	AY099809 Arabidops
33	97.6	8.2	1121	8	FB010377	AB010897 Arabidops
34	94.4	7.9	1092	8	AY081540	AY054219 Arabidops
35	94.4	7.9	1221	8	AY042887	U10277 Flaveria bi
36	94.4	7.9	1347	8	AY087493	AY042887 Arabidops
37	94.4	7.9	108387	8	AC016662	AY087493 Arabidops
C 38	92.2	7.7	44879	8	AC026238	AC016662 Arabidops
C 39	91.8	7.7	1236	8	FFRFLA3STB	AC026238 Arabidops
40	91	7.6	1161	8	FFRFLA3STA	M84136 Flaveria ch
41	89.2	7.5	1041	6	AX080460	M84135 Flaveria ch
42	87.6	7.3	978	8	AF000307	AX080460 Sequence
43	87.4	7.3	975	8	AF000306	AF000307 Brassica
C 44	87	7.3	172468	2	AP005589	AF000306 Brassica
C 45	86.2	7.2	124300	8	AP004534	AP005589 Oryza sat
						AP004534 Lotus jap

ALIGNMENTS

RESULT 1
AX338910
LOCUS AX338910
DEFINITION Sequence 15 from Patent WO0185971.
ACCESSION AX338910
VERSION AX338910.1 GI:18129074
KEYWORDS
SOURCE Zostera marina.
ORGANISM Zostera marina
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Zosteraceae; Zostera.
1
REFERENCE
AUTHORS Alberte, R.S. and Smith, R.D.
TITLE Transgenic plants incorporating traits of Zostera marina
JOURNAL Patent: WO 0185971-A 15 15-NOV-2001;

```

Phycogen, Inc. (US)
Location/Qualifiers
1. .1192
/organism="Zostera marina"
/db_xref="taxon:29655"
join(<1. .33.37. 1041.1045. 1059.1063. 1077.1081. 1086.
1090. 1119.1123. >1191)
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD20565.1"
/db_xref="GI:18129075"
/translation="TRGTTGIAVACLIPLIMAGILALEKCFGSKNEOEKEDSKMYKRY
REIVSLPNDYNDVDPSPPTHELLFFNFHSVCQVLELYLYMGRENTMPDLNLSPR
WTKALIFAILTRDVPNHPSPFTKIIINISNRKSTFVSFWKGLNLPDLKDLKSDVDF
LFAHPTPSLLPASVLSKGTKIINISNRKSTFVSFWKGLNLPDLKDLKSDVDF
ASGISFCGPEWFAETNAASTNLLLSYEEMLEKPVENVKLAEPFGCGGTFDDE
EKQIIVDEIVKLCSPFNKQVQVNGSSVNSKIDNKHFRKGEVRDMANYLTSEMIK
KLETAGKINSEKHLNKNLTNFNLNVCESKSNISLCLLSYSFVINPILKXKKK"
BASE COUNT 391 a 206 c 266 g 329 t
ORIGIN
Query Match 100.0%; Score 1192; DB 6; Length 1192;
Best Local Similarity 100.0%; Pred. No. 1.8e-264;
Matches 1192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACAGCGGAATACTGAATCGCTGTTGCTGTAGTACCTGATGATGCTGGAATT 60
DB 1 ACAGCGGAATACTGAATCGCTGTTGCTGTAGTACCTGATGATGCTGGAATT 60
QY 61 TTAGCTTTGGAGAATGTTTCGGATCCAAAGATGAGCAAGAGAGGAAGAGATTCGAA 120
DB 61 TTAGCTTTGGAGAATGTTTCGGATCCAAAGATGAGCAAGAGAGGAAGAGATTCGAA 120
QY 121 ATGTACAGAGATATAGAGAGATGTTTCTTCACTTCCCTCGAATGATTTTGGGGGAT 180
DB 121 ATGTACAGAGATATAGAGAGATGTTTCTTCACTTCCCTCGAATGATTTTGGGGGAT 180
QY 181 ACCATGAGTTGTACAGGAGATTTGGCAATGGATATCTTGACCTGATCATGGCT 240
DB 181 ACCATGAGTTGTACAGGAGATTTGGCAATGGATATCTTGACCTGATCATGGCT 240
QY 241 TTCAAGATATTTCAAGGCTCGAGAGACGACATATCTTACGACTCTTCCAAAGGCT 300
DB 241 TTCAAGATATTTCAAGGCTCGAGAGACGACATATCTTACGACTCTTCCAAAGGCT 300
QY 301 GGAACGACATGGACGAAGGCACTGAGCTTTGCCATCTTAACACAGAGATTTAACCCCA 360
DB 301 GGAACGACATGGACGAAGGCACTGAGCTTTGCCATCTTAACACAGAGATTTAACCCCA 360
QY 361 TCATCACCAGACATCCACTTCTTCTTCAACCTCATCTGCTGTTTCAAAATTTGGAG 420
DB 361 TCATCACCAGACATCCACTTCTTCTTCAACCTCATCTGCTGTTTCAAAATTTGGAG 420
QY 421 TATTGTACATGGGTAGAGAAATACGATGCCAGCTCGATATGTTGAATGAATGCCCG 480
DB 421 TATTGTACATGGGTAGAGAAATACGATGCCAGCTCGATATGTTGAATGAATGCCCG 480
QY 481 AGGTTGTTGGCGGACATCCACTACTCTTTGTTGCCGGCTCTGTTTCAAAATTCGGA 540
DB 481 AGGTTGTTGGCGGACATCCACTACTCTTTGTTGCCGGCTCTGTTTCAAAATTCGGA 540
QY 541 ACAAAATCATCAATATAGCGCAACCGTAAAGTACATTTGCTCTTTTGGAAATTT 600
DB 541 ACAAAATCATCAATATAGCGCAACCGTAAAGTACATTTGCTCTTTTGGAAATTT 600
QY 601 GGCAATCTGATTAACCCGACAAAGTTATTTGACCTCGAAAGAGCGTTGATCTTCGCA 660
DB 601 GGCAATCTGATTAACCCGACAAAGTTATTTGACCTCGAAAGAGCGTTGATCTTCGCA 660
QY 661 TCGGGAATCTCTTTTGTGGACCGAATGGAATTTCCAGCGGAGCTTCCCAATCGCGCG 720
DB 661 TCGGGAATCTCTTTTGTGGACCGAATGGAATTTCCAGCGGAGCTTCCCAATCGCGCG 720

```

Phycogen, Inc. (US)
Location/Qualifiers
1. .1192
/organism="Zostera marina"
/db_xref="taxon:29655"
join(<1. .33.37. 1041.1045. 1059.1063. 1077.1081. 1086.
1090. 1119.1123. >1191)
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD20565.1"
/db_xref="GI:18129075"
/translation="TRGTTGIAVACLIPLIMAGILALEKCFGSKNEOEKEDSKMYKRY
REIVSLPNDYNDVDPSPPTHELLFFNFHSVCQVLELYLYMGRENTMPDLNLSPR
WTKALIFAILTRDVPNHPSPFTKIIINISNRKSTFVSFWKGLNLPDLKDLKSDVDF
LFAHPTPSLLPASVLSKGTKIINISNRKSTFVSFWKGLNLPDLKDLKSDVDF
ASGISFCGPEWFAETNAASTNLLLSYEEMLEKPVENVKLAEPFGCGGTFDDE
EKQIIVDEIVKLCSPFNKQVQVNGSSVNSKIDNKHFRKGEVRDMANYLTSEMIK
KLETAGKINSEKHLNKNLTNFNLNVCESKSNISLCLLSYSFVINPILKXKKK"

BASE COUNT 391 a 206 c 266 g 329 t

ORIGIN

Query Match 100.0%; Score 1192; DB 6; Length 1192;
Best Local Similarity 100.0%; Pred. No. 1.8e-264;
Matches 1192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGCGGAATACTGAATCGCTGTTGCTGTAGTACCTGATGATGCTGGAATT 60
DB 1 ACAGCGGAATACTGAATCGCTGTTGCTGTAGTACCTGATGATGCTGGAATT 60

QY 61 TTAGCTTTGGAGAATGTTTCGGATCCAAAGATGAGCAAGAGAGGAAGAGATTCGAA 120
DB 61 TTAGCTTTGGAGAATGTTTCGGATCCAAAGATGAGCAAGAGAGGAAGAGATTCGAA 120

QY 121 ATGTACAGAGATATAGAGAGATGTTTCTTCACTTCCCTCGAATGATTTTGGGGGAT 180
DB 121 ATGTACAGAGATATAGAGAGATGTTTCTTCACTTCCCTCGAATGATTTTGGGGGAT 180

QY 181 ACCATGAGTTGTACAGGAGATTTGGCAATGGATATCTTGACCTGATCATGGCT 240
DB 181 ACCATGAGTTGTACAGGAGATTTGGCAATGGATATCTTGACCTGATCATGGCT 240

QY 241 TTCAAGATATTTCAAGGCTCGAGAGACGACATATCTTACGACTCTTCCAAAGGCT 300
DB 241 TTCAAGATATTTCAAGGCTCGAGAGACGACATATCTTACGACTCTTCCAAAGGCT 300

QY 301 GGAACGACATGGACGAAGGCACTGAGCTTTGCCATCTTAACACAGAGATTTAACCCCA 360
DB 301 GGAACGACATGGACGAAGGCACTGAGCTTTGCCATCTTAACACAGAGATTTAACCCCA 360

QY 361 TCATCACCAGACATCCACTTCTTCTTCAACCTCATCTGCTGTTTCAAAATTTGGAG 420
DB 361 TCATCACCAGACATCCACTTCTTCTTCAACCTCATCTGCTGTTTCAAAATTTGGAG 420

QY 421 TATTGTACATGGGTAGAGAAATACGATGCCAGCTCGATATGTTGAATGAATGCCCG 480
DB 421 TATTGTACATGGGTAGAGAAATACGATGCCAGCTCGATATGTTGAATGAATGCCCG 480

QY 481 AGGTTGTTGGCGGACATCCACTACTCTTTGTTGCCGGCTCTGTTTCAAAATTCGGA 540
DB 481 AGGTTGTTGGCGGACATCCACTACTCTTTGTTGCCGGCTCTGTTTCAAAATTCGGA 540

QY 541 ACAAAATCATCAATATAGCGCAACCGTAAAGTACATTTGCTCTTTTGGAAATTT 600
DB 541 ACAAAATCATCAATATAGCGCAACCGTAAAGTACATTTGCTCTTTTGGAAATTT 600

QY 601 GGCAATCTGATTAACCCGACAAAGTTATTTGACCTCGAAAGAGCGTTGATCTTCGCA 660
DB 601 GGCAATCTGATTAACCCGACAAAGTTATTTGACCTCGAAAGAGCGTTGATCTTCGCA 660

QY 661 TCGGGAATCTCTTTTGTGGACCGAATGGAATTTCCAGCGGAGCTTCCCAATCGCGCG 720
DB 661 TCGGGAATCTCTTTTGTGGACCGAATGGAATTTCCAGCGGAGCTTCCCAATCGCGCG 720

721 TCTACTAATCAAACTTCTGCTATCTTCTGAGTTACGAGAAATGTTAGAGAACCCAGTTGAA 780
721 TCTACTAATCAAACTTCTGCTATCTTCTGAGTTACGAGAAATGTTAGAGAACCCAGTTGAA 780
781 AATGTGAAGAGCTAGCTGAGTTTCATGGGATGTTGGGTTTCAGACGATGAGAGAACAA 840
781 AATGTGAAGAGCTAGCTGAGTTTCATGGGATGTTGGGTTTCAGACGATGAGAGAACAA 840
841 GGGATTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
841 GGGATTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
901 AACAAAAACCGATCAAGTCAATTCGAAAAACGAAATCGAACAAAGCATTTCTTCAGGAAAGT 960
901 AACAAAAACCGATCAAGTCAATTCGAAAAACGAAATCGAACAAAGCATTTCTTCAGGAAAGT 960
961 GAGGTGAGAGATTGGGCAAACTATCTACCTCGGAAATGATTAGAAACTCGAGACGCC 1020
961 GAGGTGAGAGATTGGGCAAACTATCTAACCTCGGAAATGATTAGAAACTCGAGACGCC 1020
1021 GGAAAAATAAATGAATCAGAGTAAAGCATTTATTATCGTGAATAAAGAAATCTTACATGA 1080
1021 GGAAAAATAAATGAATCAGAGTAAAGCATTTATTATCGTGAATAAAGAAATCTTACATGA 1080
1081 AACTTCTGAAATCTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140
1081 AACTTCTGAAATCTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140
1141 TCGTATTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1192
1141 TCGTATTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1192

RESULT 2
ATT14D3/c 88010 bp DNA linear PLN 04-PEB-2000
LOCUS Arabidopsis thaliana DNA chromosome 3, BAC clone T14D3.
DEFINITION AL138649
ACCESSION AL138649
VERSION AL138649.1 GI:6899911
KEYWORDS Arabidopsis thaliana.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 88010)
Jordan, N., Bangert, S., Wiedelmann, R., Voss, H., Unseld, M.,
Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetier, F. and Salanoubat, M.
Unpublished
2 (bases 1 to 88010)
EU Arabidopsis sequencing, project.
Direct Submission
Submitted (04-PEB-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project
Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
Gaston Cremieux, Bp191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES
source
1. .88010
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
1145. .2116
/gene="T14D3.10"
complement(1145. .2116)
/note="strong similarity to FLAVONOL 4'-SULFOTRANSFERASE -
Flaveria chloraefolia, EMBL:M84136"

gene
CDS


```
exon      21360..21450
           /gene="T14D3.40"
           /number=7
intron    21451..21532
           /gene="T14D3.40"
           /number=7
exon      21533..21625
           /gene="T14D3.40"
           /number=8
gene       22937..23299
           /gene="T14D3.50"
           /number=7
           /gene="T14D3.50"
           /codon_start=1
           /product="putative protein"
           /protein_id="CAB72149.1"

Query Match      12.2%; Score 145.4; DB 8; Length 88010;
Best Local Similarity 51.4%; Pred. No. 3.7e-23;
Matches 528; Conservative 0; Mismatches 466; Indels 33; Gaps 7;

QY 15 TGAATCGCTGTTGCTTGTAGCTACACCTGATAATGGCTGGAAATTTTACCTTTGGAGAA 74
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 75 ATGTTTCGGATCCAGAAATGAGCAAGAGAGAGAGATTCCAAAATGTACAAAGAGATA 134
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2144 ATCAATTTGTAATGATGTGAAGAAGATGGAGATGAATCTGAGAAATGGAACCTTAA 2085
QY 135 TAGAGAGA-----TTGTTTCTTCACTTCCCTC---GAATGATTATTTGGGGGATAC 182
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
PB 2084 CGAAGAAACAAGACTCTAATCTCTTCACTTCTTCAAGACAAGATTTCACCTGGGAAAC 2025
QY 183 CATGAGTTGTACAAGGATTTGGCAATGGGATATCTTGTACCTGGTATCATGCGTTT 242
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2024 TATTGCAAGTATCAAGGATGTTGGTATCTCAACAATGTTCTTCAAGCTGTCTCAATTT 1965
QY 243 CGRAGATAATTTCAAGGCTCGAGAGCGGACATTATCTTACGACTCTCCAAAGGCTGG 302
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1964 CCAGAAAGCTTCAAGCCTCAAGACACGATATCATGTTGCTTCCTTAATACGCG 1905
QY 303 AACGACATGGACGAGGACCTGAGCTTTGGCCATCTTAACACGAGATGTTAACCCACATC 362
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1904 CACCACCTGGCTCAAGGCGCTTACATTCGCACTCTTATAGATCAAAACAGCCTTCTCA 1845
QY 363 ATCAGGACACATCCACTTTGTTCTTCAACCTCATTCGTTGCTTCAAAATTTGGAGT- 421
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1844 TGATGATGATCATCTCTTCTTCTTAATATCCACAGTTCTTGTACCTTACTTTGAGAT 1785
QY 422 --ATTGTACATGGTAGAGAAATACGATGCCAGACCTCGATATGTTGAATGAATCGCC 479
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1784 AGATCTCTATCTAGTACGAAAT-----CTTGACCTTACCAAGTTCTCATCTCC 1731
QY 480 GAGGTTGTTTCGGACACATCCCCATCTTTTGTTCGGCGCTGTGTTTGAATCGG 539
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1730 GAGGCTGTTTTCGACACAGTGCCTGCATACATGTTGCAAGAAGTCTCAAGGTTCTAC 1671
QY 540 AACAAAATCATATATAAGCCGCAACCGTAAGAGTACATTGCTCTTTTTCGAAT 599
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1670 TTGTAAATTTGTATATATCTAGAAACGTAAGACACATTTGTTTCTATTTGTCATTT 1611
QY 600 TGGCAATCTGATTAAACCCCG-----ACAAGTTATTGGACCTCGAAAAGAGCGTTGAT 653
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1610 CTTTACTAAGAACAAACCGATGAANAATAATTAAGCATGTTTCGAGGATACGTTTGAT 1551
QY 654 CTTCCGATCGGAATCTCTCTTTTGTGGACCGAATGGAATTTCCAGCGGAGTTCACCA 713
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1550 GTTTTGTAGGAGTACGATTTTCGGGCTTTTTCGGATCATGCTTTAAGCATTTGGAG 1491
QY 714 TCGGGCGTCTACTAATTCAAAC---TTGCTATTGTTGAGTTACGAGAATAATGTTAGAA 770
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1490 AGGAAGCTTGGAGATCCAAACCATGCTGCTTTTATGAAGTTTGAAGATGAAGACGA 1431
```

```
QY 771 GCCAGTTGAAAATGTGAAGAGCTAGCTGAGTTCATGGATGTGGTTTCACAGACGATGA 830
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1430 ACCTCGTACCAGATCAAGAAATTTGCGAGTTCCTTAGGTTGCTCTTTTACTAAGGAGA 1371
QY 831 GGAGAACACAGGATTTGTTGATGAGATAGTTAAACTTTTGTAGCTTCGACAACTCTGAAGAA 890
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1370 AGAAGAGAGCGGATCGGTGGATGAGATTATCGATCTTTTGTCTCTACGTATCTGAGCAG 1311
QY 891 TCAACAGGTGAACAAAACCGGATCAAGCTACAATTCGAAAATCGACAAACAAGCAATTTCTT 950
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1310 TTTGGAGATCAATAGACCGG---AAATTTGAATCTGGTAGAGAAAACAAAATGCTTTT 1254
QY 951 CAGGAAAGGTGAGTGAGAGATTGGGCAAACTATCTTAACGTGCGAAATGATTAGAAACT 1010
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1253 CGGTAAAGGAGAGTTGTTGATTGGAAGAACTATTGTAATCTCTGAAATGGAGAACAAAAT 1194
QY 1011 GGAGAGC 1017
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1193 CGACATG 1187

RESULT 3
ATF14D17/c
LOCUS      89214 bp      DNA      linear      PLN 28-APR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone F14D17.
ACCESSION  AL353992
VERSION     AL353992.1  GI:7671394
KEYWORDS   Arabidopsis thaliana.
SOURCE     Arabidopsis thaliana.
ORGANISM   Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.

REFERENCE  1 (bases 1 to 89214)
AUTHORS   Jordan,N., Bangert,S., Wiedelmann,R., Voss,H., Unseld,M.,
Mewes,H.W., Rudd,S., Lemcke,K., Mayer,K.F.X., Quetier,F. and
Salanoubat,N.
Unpublished
2 (bases 1 to 89214)
EU Arabidopsis sequencing,project.
Direct Submission
Submitted (28-APR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,Project
Coordinator: Marcel Salanoubat and Francis Quetier, Grouperment
d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue
Gaston Cremieux, Bp191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.

FEATURES
source
    1..89214
        /organism="Arabidopsis thaliana"
        /variety="Columbia"
        /db_xref="taxon:3702"
        /chromosome="3"
        5576..6238
        /gene="F14D17.10"
        complement(5576..6238)
        /gene="F14D17.10"
        complement(5576..6238)
        /gene="F14D17.10"
        /notes="similarity to predicted protein, Arabidopsis
        thaliana"
        /codon_start=1
        /product="putative protein"
        /protein_id="CAB89409.1"
        /db_xref="GI:7671395"
        /translation="MELEQTKLVAHELKRDQLIHLEVLTKTKERDEALEKYNH
        LLNLLQLQQQNNQKQELVTPPLSGASSIIEDEQVQPPQNLNSKFSSTDE
        ELSMSPSVIDPVTNNQOIEVSGDEMLATLLPDKPLPKGLQAVKAGPLLOTLLA
        GPLQWRHPPPPLETSIPVPTIPLQFQNNCGNSNKKRAFSISDEYSETKYQVL
```

```

exon      LH
complement(5576..6238)
/gene="F14D17_10"
/number=1
/gene="F14D17_20"
join(12760..12885,12909..13001)
/gene="F14D17_20"
/codon_start=1
/product="putative protein"
/protein_id="CAB89310.1"
/db_xref="GI:7671396"
/translation="MTLAVVAPTTVVASSGGGCGGSGDNGSGDNGNPNNGVAVPG
NSGGGGDDGGVGGDDGGGGVWLW"
12760..12885
/gene="F14D17_20"
/number=1
12886..12908
/gene="F14D17_20"
/number=1
12909..13001
/gene="F14D17_20"
/number=2
18756..18827
/product="tRNA-Gly"
/notes="tRNA predict as a tRNA- Gly : anticodon tcc"
19510..20866
/gene="F14D17_30"
join(19510..19614,20074..20142,20264..20387,20469..20529,
20602..20866)
/gene="F14D17_30"
/codon_start=1
/product="putative protein"
/protein_id="CAB89311.1"
/db_xref="GI:7671397"
/translation="MRQSLSDITNSQSQBELNLHKISBPSPGHVRLMKVKALHHEVT
CKNALKAKCCREKENTKPRNALITVESVLNITDEDSPPKPFVPTRRFRIRSKSLG
ASTTNKEAKVKESKRRHLRRSPVRSAAQEVTFENLFEIEDLQLTMPDNMSQQDN
19510..19614
/gene="F14D17_30"
/number=1
19615..20073
/gene="F14D17_30"
/number=1
20074..20142
/gene="F14D17_30"
/number=2
20143..20263
/gene="F14D17_30"
/number=2
20264..20387
/gene="F14D17_30"
/number=3
20388..20468
/gene="F14D17_30"
/number=3
20469..20529
/gene="F14D17_30"
/number=4
20530..20601
/gene="F14D17_30"
/number=4
20602..20866
/gene="F14D17_30"
/number=5
24158..26563
/gene="F14D17_40"
join(24158..24366,24453..24777,24884..25036,25129..25377,
25629..25721,25837..25843,25968..26074,26339..26563)
/gene="F14D17_40"
/notes="strong similarity to cytochrome P450 dl3695c,
Arabidopsis thaliana, PIR:C71417"

/codon_start=1
/product="cytochrome P450-like protein"
/protein_id="CAB89312.1"
/db_xref="GI:7671398"
/translation="MNLFWNTGFCVILVVARVGHWWYQMSNPKSNGKLPFGSMGFPPI
IGETLDFPKPYGFEISPYLKKKMLRYGFLFRNLILGVTVVSTDKVNMELRQENK
SFILSYDPGLMKPLGKDSFLKIGNIHKIQTITLLSSEGLKRIKIDMDRVTREH
LSKAKTGRLDVDAVSKLI IAHLTPKMSNLKPQOAKLMTGIFKAPTDFMPTSYLI
SAGKGLYNTLWACREGMREIKDIYTMKRTSEKYYDDPLNATAIEESKAGELNENAI
TLITLSCVTODTTSKAI CLAVKFLLENPKVLAELKKEHEVILESEDREKGGVTWEY
RKMTFTNKSQGTTPAGWIVMLIPSVVHFDEIYENPEFNPWRWEGKELRAGSKTF
WFGTLRQCAGAEFARLOISVFLHHLVTTYNFSLHODCEVLVPAALHPNGISINIS
KCPK"
24158..24366
/gene="F14D17_40"
/number=1
24367..24452
/gene="F14D17_40"
/number=1
24453..24777
/gene="F14D17_40"
/number=2
24778..24883
/gene="F14D17_40"
/number=2
24884..25036
/gene="F14D17_40"
/number=3
25037..25128
/gene="F14D17_40"
/number=3
25129..25377
/gene="F14D17_40"
/number=4
25378..25628
/gene="F14D17_40"
/number=4
25629..25721
/gene="F14D17_40"
/number=5
25722..25836
/gene="F14D17_40"
/number=5
25837..25843
/gene="F14D17_40"
/number=6
25844..25967
/gene="F14D17_40"
/number=6
25968..26074
/gene="F14D17_40"
/number=7
26075..26338
/gene="F14D17_40"
/number=7
26339..26563
/gene="F14D17_40"
/number=8
32738..33813
/gene="F14D17_50"
join(32738..32746,32854..32996,33276..33354,33471..33624,
33686..33813)
/gene="F14D17_50"
/codon_start=1
/product="putative protein"
/protein_id="CAB89313.1"
/db_xref="GI:7671399"
/translation="MAELGSSGKSTLASIANKMVQAFKIEETDFFLLSSPNSYKDSI
DYTSKLREARLRWRDTRATAADICHIARVKGDKSQNGKIEPKSKRKHLLSLNR
KWASNEURRALLQRGDSIAEKSSIGKRDSSRYREVPTPASFSSTGVGGEIRDSFF
DNSVRELNGY"
32738..32746
/gene="F14D17_50"

```

intron		/number=1 32747...32853 /gene="F14D17_50" /number=1
Query Match		12.2%; Score 145.4; DB 8; Length 89214;
Best Local Similarity		51.4%; Pred. No. 3.7e-23;
Matches		528; Conservative 0; Mismatches 466; Indels 33; Gaps 7;
QY	15	TGGAATCGCTGTTGCTGTAGTACCTACCTGATTAATGGCTGGTAATTTAGCTTTTGAGAA 74
DB	79152	TGAAACCGTGCATTTTGCACAAACATACAAATCATAGTGATGTATATATAGTAGA 79093
QY	75	ATGTTTCGGATCCAAGATCAGCAAGAGAGGAGAGATTCACAAATGTACAAAGAGATA 134
DB	79092	ATCATTTGTTGTAATGATGTGAAGAAGATGGAGATGAATCTGAGAAATGAAGACCTTAA 79033
QY	135	TAGAGAGA-----TTGTTTCTTCACCTCCCTC---GAATGATATTGTTGGGGGATAC 182
DB	79032	CGAAGAAACAAAGACTTAATCTCTTCACTTCTTCAGACAAAGATTCACCTGGGAAAC 78973
QY	183	CATGAGTTGTACAAGGATTTGGCAATGGGATATCTGTACCTGGTATCATGCTTT 242
DB	78972	TATTTGCAAGTATCAAGGATGTTGGTATACTCAATGTTCTTCAAGCTGTCTCAATTT 78913
QY	243	CGAAGATTAATTTCAAGCTCGAGAGCGGACATTATCTTACGACTCTTCCAAAGGCTGG 302
DB	78912	CCAGAAAGCTTCAGCCTCAAGACACCGATATCATGTTGCTTCCTTAATGCGG 78853
QY	303	AACGACATGACGAAGACATGCTGTTTGCATCTTAAACAGAGATGTTAAACCAACCATC 362
PB	78852	CACCACCTGGCTCAAGCGCTTACATTCGCACTCCITCATAGATCAAAACAGCCTTCTCA 78793
QY	363	ATACCGACACATCCACTTTTGTCTTCAACCTCATCTGCTGTTCAAAATTTGGAGT- 421
DB	78792	TGATGATGATCATCTCTCTTCTTAATAATCCACAGCTTCTTGTACCTACTTTGAGAT 78733
QY	422	--ATTGTACATGGGTAGAGAAATACGATCCAGACCTCGATATGTTGAATGAATCGCC 479
DB	78732	AGATCTATCTACGTAGCGAAT-----CTGACCTTACCAGTTCTCATCACTCC 78679
QY	480	GAGTGTGTTGCGGACACATCCCATACTCTTTGTCGCGGCTGCTGTTTGAATCGGG 539
DB	78678	GAGCTGTTTTTCGACACACGTCGCTCACATAGCTTGCAAGAAGTCTCAAGGTTCTAC 78619
QY	540	AACAAATCATCAATATAGCCGACCGCTAGAGTACATTTGTGCTTTTGGAAATT 599
DB	78618	TTGTAAATTTGTGTATATCTAGAAACGTAAAGACACATTTGGTTTCATATTGGCATTT 78559
QY	600	TGCAATCTGATTAACCCCG-----ACAAGTTATTGGACCTCGAAAGAGCGTTGATAT 653
DB	78558	CTTTACTAAGAAACAACCCGATGAAATAATAAGCAGTTTCGAGGATACGTTTGAGAT 78499
QY	654	CTTCGCATCGGAAATCTCTTTTGTGACCGGAATGGAATTTCCAACGGAGTTTCAACAA 713
DB	78498	GTTTGTAGGGAGTACGATTTTCGGGCTTTTGGGATCATGTCTTTAAGCTATTGGAG 78439
QY	714	TGCGGCTCTACTAATTCAAAC---TTGCTATTGTTGAGTTACGAAAGATTTAGAGAA 770
DB	78438	AGGAAGCTTGAAGATCCAACCATGTCCTTTTATGAAGTTTGAAGATGAAAGCAGA 78379
QY	771	GCCAGTTTGAAATGTGAAGAGCTAGCTGATGTTTCATGGGATGTTGCGTTCACAGACGATGA 830
DB	78378	ACCTCGTGACAGATCAAGAAATTTGCGAGTTCTTAGGTTGCTTTTACTAAGGAAGA 78319
QY	831	GGAGAAACAAGGATTTGTTGATGATAGATAGTTAACTTTGTAGCTTCGACAAATCTGAAGAA 890
DB	78318	AGAGAGAGCGGATCGGTGATGATGATATTCGATCTTTGTTCTCTACGTAATCTGACGAG 78259
QY	891	TCAACAGGTGAACAAAACGATCAAGCTACAAATTCGAAATTCGAAACCAAGCATTTCTT 950
DB	78258	TTTGAGATCATTAAGACCGG-----AAATTTGATTTGTTAGAGAAACAAATGTTTTT 78202
RESULT 4		
AC119410		
LOCUS		
DEFINITION		Medicago truncatula clone mth2-13n10, WORKING DRAFT SEQUENCE, 14
unordered pieces.		
AC119410		
VERSION		AC119410.1 GI:20330840
HTG; HTGS PHASE1; HTGS_DRAFT.		
KEYWORDS		Medicago truncatula.
SOURCE		Medicago truncatula
ORGANISM		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE		1 (bases 1 to 98198)
AUTHORS		Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE		Medicago truncatula BAC Clone mth2-13n10
JOURNAL		Unpublished
REFERENCE		2 (bases 1 to 98198)
AUTHORS		Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE		Direct Submission
JOURNAL		Submitted (26-APR-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
COMMENT		----- Genome Center Center: Department Of Chemistry And Biochemistry The University Of Oklahoma Center code:UOKNOR ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 14 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 2444: contig of 2444 bp in length * 2445 2544: gap of unknown length * 2545 5232: contig of 2688 bp in length * 5232 5333: gap of unknown length * 5333 8626: contig of 3294 bp in length * 8626 8727: gap of unknown length * 8727 12541: contig of 3815 bp in length * 12541 12642: gap of unknown length * 12642 16370: contig of 3729 bp in length * 16370 16470: gap of unknown length * 16470 22893: contig of 6423 bp in length * 22893 22994: gap of unknown length * 22994 30423: contig of 7430 bp in length * 30423 30523: gap of unknown length * 30523 36817: contig of 6294 bp in length * 36817 36918: gap of unknown length * 36918 45199: contig of 8281 bp in length * 45199 45299: gap of unknown length * 45299 52079: contig of 6681 bp in length * 52079 60174: contig of 8095 bp in length * 60174 60275: gap of unknown length * 60275 66818: contig of 6544 bp in length * 66818 66919: gap of unknown length * 66919 81725: contig of 14807 bp in length

FEATURES	81726	81825:	gap of unknown length	
SOURCE	81826	98198:	Contig of 16373 bp in length.	
	1.	98198	Location/Qualifiers	
		/organism="Medicago truncatula"		
		/db_xref="taxon:3880"		
		/clone="mth2-13n10"		
		/clone.lib="Medicago truncatula BAC library H2"		
BASE COUNT	31816	a	16153 c 16811 g 32014 t 1404 others	
ORIGIN				
Query Match	11.0%; Score 131.4; DB 2; Length 98198;			
Best Local Similarity	50.2%; Pred. No. 6.2e-20;			
Matches 442; Conservative	0; Mismatches 421; Indels 18; Gaps 4;			
Qy	190	TTGTCAACAGGATTTTGGCAATGGGATATCTTGTACCTGGTATCATGGCTTTTCGAAGAT	249	
Db	2709	TTTATCATGGATTTTGGTGTCCATCACTCTTATCCATCTGTGAACCTTTTCAAAAT	2768	
Qy	250	AATTTCAAGGCTCGAGAGCGGACATTTATCTTTACGACTTTCCAAAGGCTGGAACGACA	309	
Db	2769	AATTTTCATGCCAAGATAGTAGTATTTGTAGTTGCATCCATGCCAATACTGGCACTACT	2828	
Qy	310	TGACGAGGACCTAGCTTTGGCATCTTCAACGAGATGTTAAACCCATCATCACCG	369	
Db	2829	TGGCTTAAAGGCTTGCATATGCTATTGTTAATCGCCAACTTTTACCTCTTAGAGAAT	2888	
Qy	370	ACATCTCCATTTTGTCTTCAACCTCATCTGTGTGTTCAAAATTTGGAGTATTTGTAC	429	
Db	2889	AACATCCATTACTATCAITTAATCACTGAACTTTGTGCTCATTTTGAAGTAAGTGT	2948	
Qy	430	ATGGGTAGAGAAATACGATGCGACAGCTCGATATGTTGAAT---GAATCGCGAGGTTG	486	
Db	2949	AATAATACAGATGTCATGCTCAATGCTCAATTTGATGTACCAATATGTTGAACCAAGCTT	3008	
Qy	487	TTTGGCGGACACATCCCATACTCTTTGTGTCGGGCTGTCTTTTGAATTCGGGAATTTG	546	
Db	3009	TATGGAACTCACATGCTCATCTTCTTGTGTCGCAAGTCAATTCAGAGTCCAAATTTGTA	3068	
Qy	547	ATCATCAATTAAGCGCAACCGTAAGAGTACATTTGTGTCTTTTGGAAATTTGGCAA-	605	
Db	3069	ATAATTTATATATGAGAACCCATTTGATGATATTTGATATATTTGACATTTTCATCAAC	3128	
Qy	606	-----TCTGATTAACCCGCAAGTATTGGAACCTCGAAAGAGCGTTGATATCTTC	657	
Db	3129	AAACTCAGATTAAAGAAATCTTTAACTGAATTAACTTTGAGGAATCTTTTGAAGGTAC	3188	
Qy	658	GCATCGGAATCTCTTTTGTGACCGGAATGGAATTTCCAAAGCGGAGTTCCCAATGCG	717	
Db	3189	TGTAAGGGAATATGTTTTTTTGGTCCGCTCTGGGCAATATGTTGGGTTACTTTGAAGGAG	3248	
Qy	718	CGGTCTACTAATCA---AACTGCTATTCTGTAGTTACGAAGAAATGTTAGAGAAGCCA	774	
Db	3249	AGCATAGAAACACAGATAGGTTTATCTTGAATATGAAGACCTTAAAGAAAGATGTT	3308	
Qy	775	GTTGAAAATGTGAAGAAGCTAGCTGAGTTTCATGGATGTGGGTTTCACAGACGATGAGGAG	834	
Db	3309	AATTTTCACAGAAAGAAATTCGAGAAATTTGTGGATTTCTTCTCACTGAAGAAGGAA	3368	
Qy	835	AAACAAGGATTTGTGATAGATAGTTTAAATTTGTAGTTTCGACAAATCTGAAGATCAA	894	
Db	3369	AATAATGGAGTATTGAAAATATAATCAAGTTATGCACTTTGAGAGTATGAAGGAATCA	3428	
Qy	895	CAGTGTGAACAAAACGATCAAGCTACAATTCGAAATCCACACAGACATTTCTTCAGG	954	
Db	3429	AGTAAAATAAATCTGGAACCTGCAGCACTA---AAGATTGAGAGAGATTTTCTTCGA	3485	
Qy	955	AAAGGTGAGGTGAGATTTGGGCAAACTATCTAAGTCGGAATATGTTAAGAAACTGGAG	1014	
Db	3486	AAGGAGAAATAGGGATTTGGTGAATATCTCTCCCTTCGATGATAAAAAAATATATCC	3545	
Qy	1015	ACGCCCGGAAAAATPAAATGAATCAAGTAAAGCAATTTATT	1055	

Db	3546	AAAGTCATGGAAGAAAAAATAAGTGGATCAAGCCATCATTT	3586	
RESULT 5				
AC027134				
LOCUS				
DEFINITION				
AC027134				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				

Arabis thaliana chromosome I BAC F13B4 genomic sequence, complete sequence.

AC027134.4 GI:8576187

HTG

Arabidopsis thaliana

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 102183)

Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Miranda, M., Brooks, S., Buehler, E., Chou, J., Choi, E., Dunn, P., Gonzalez, A., Howing, B., Kim, C., Koo, T., Lee, J.M., Lenz, C., Liu, A., Liu, S., Mukharzky, N., Pham, P., Sakano, H., Shinn, P., Toriumi, M., Vaysberg, M., Walker, M., Yu, G., Ecker, J., Theologis, A., and Davis, R.W.

Unpublished

2 (bases 1 to 102183)

Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Bei, O., Buehler, E., Chin, C., Chou, J., Choi, E., Dunn, P., Gonzalez, A., Howing, B., Kim, C., Koo, T., Lee, J.M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharzky, N., Pham, P., Sakano, H., Schwartz, J., Shinn, P., Thaveri, A., Toriumi, M., Vaysberg, M., Walker, M., Yu, G., Ecker, J., Theologis, A., and Davis, R.W.

Direct Submission

Submitted (28-MAR-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

3 (bases 1 to 102183)

Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Bei, O., Buehler, E., Chin, C., Chou, J., Choi, E., Dunn, P., Gonzalez, A., Howing, B., Kim, C., Koo, T., Lee, J.M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharzky, N., Pham, P., Sakano, H., Schwartz, J., Shinn, P., Thaveri, A., Toriumi, M., Vaysberg, M., Walker, M., Yu, G., Ecker, J., Theologis, A., and Davis, R.W.

Direct Submission

Submitted (21-JUN-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

4 (bases 1 to 102183)

Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Ecker, J., Theologis, A., and Davis, R.W.

Direct Submission

Submitted (15-AUG-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

On Jun 21, 2000 this sequence version replaced gi:8134868.

Bases 1-55,850 of F13B4 overlap with bases 47,308-103,157 of BAC clone T6J4, gb|AC011810 and bases 76,686-102,183 of F13B4 overlap with bases 1-25,498 of BAC clone F21F23, gb|AC027656.

e-mail for correspondence: arab@sequence.stanford.edu

Genes with similarity to proteins in the databases are named 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The gene prediction programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge, <http://genes.mit.edu/GENSCAN/>), FEXA (Victor Solovyev, <http://genomic.sanger.ac.uk/gf/gf.shtml>), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of

Query Match	88	
Best Local		35704
Matches 45		
Qy		148
Db		35764
Qy		208

35824 TACGACAAAGATATCTTCCAAGCAATCTCAATTTCAACAAAACCTTTTCAGCCCAAGAA 35883
QY 268 ACCGACATATCTTTCAGACTCTTCCAAAGGCTGGAACGACATGGAAGGACGACTGACG 327
35884 ACGGATATATTTGCTCTTTCCTCCCAATCGGATACGACTTGGCTCAAGGCACTCA 35943
QY 328 TTGCGCATCTTAACACGAGATGTTAAACCCATCATCCGACACATCCACTTTTGTTC 387
Db 35944 TTGCGCATCGCGCAAGA-----TCAAAACACACTTCAGACAATCATCTCTGCTAACT 35997
QY 388 TTCAACCCCTATTCGCTGTTTCAAAATTTGGAGATTTGTACATGCGGTAGAGAAATACG 447
Db 35998 CATAATCCTCATGAGTAGTGCCTGCTACCTCGAGCTCGATCTTATCTCAAAAGCTCGAA 36057
QY 448 ATCCGAGACCTCGATATGTTGAATGATCGCCGAGTGTGTTGCGCGACACATCCCATAC 507
Db 36058 CCGATTTGACCAAGTTGCCATCATCTCCGAGATTTGTTCAACCCACATGCTCTT 36117
QY 508 TCTTTGTTGCGCGCTCTGTTTGAATCGGGAACAAAATCATCAATATAAGCCGCAAC 567
Db 36118 GATGCGCTTAAAGTACGTTTGAAGAGTCTCTTGAAGATCGTGACGTGTGTCAGGAAC 36177
QY 568 CGTAAGAGTACATTTGCTCTTTTGAATTTGGCAATCTGATTAACCCCGCAAGTTA 627
Db 36178 GTGAAAGACGATTTGCTATCACTTTGCTGTTTTCGAAACCTCCATGAGTGGAGAAACAAT 36237
QY 628 TTGACCTCGAAAGAGCTGTATATCTTCGCATCGGATCTCTTTTGTGACCGGAA 687
Db 36238 TTAAGTCTCGAGGCTTTGTTGAGTCTTTATGATGCGGAGTTAACTTATGCGGTCCCTTA 36297
QY 688 TGG---AATTTCAAGCGGAGTTTCAACATCGCGCTCTACTAATTCAAACTTGCTATTG 744
Db 36298 TGGGAAATGTTGAGCTATTGAGAGGAGCTTGAAGATCTTAAGCATGTGCTTTTC 36357
QY 745 TTGAGTTACGAAGAAATTTAGAGAGCCAGTTGAAATGTGAAGAGCTAGCTGAGTTC 804
Db 36358 TTGAGGTACGAGGAGTTGAAGAGCGGAGCTCTGTTGCAATCAAGAGACTTGCAGAGTTC 36417
QY 805 ATGGGATGCGGTTTCAAGAGCTAGGAGAAACAGGGATCTTGTGATGATAGTATTAA 864
Db 36418 TTGATTTGCTTCAAGAACTTAGCGTTTGGAGATCAACAAACAGG---AAGCTTGTG 36477
QY 865 CTTTGTAGCTTCGACATCTGAAGATCAACAGTGAACAAAACGATCAAGCTACAAT 924
Db 36478 CTTTGTCTTCAAGAACTTAGCGTTTGGAGATCAACAAACAGG---AAGCTTGTG 36534
QY 925 TCGAAATTCGACAAACAGCAATTTTTCAGGAAAGGTGAGGTGAGAGATTTGGGCAAACTAT 984
Db 36535 GAAGGAGTAAGTTTCAAGAGTTTTCGTAAGGGAAGTGTGATGATGGAAGAGTTAT 36594
QY 985 CTAACGTCGGAATGATTAGAACTCGAGACG 1017
Db 36595 ATGACTCCTGAATGGAACAAAATCGCATG 36627

RESULT 6
AC011810
LOCUS AC011810 103157 bp DNA linear PLN 01-SEP-2000
DEFINITION Arabidopsis thaliana chromosome 1 BAC T6J4 genomic sequence,
complete sequence.
AC011810
VERSION AC011810.8 GI:8576186
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 103157)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Howng,B., Kim,C.,
Buehler,E., Chou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
Koo,T., Lee,J.M., Lenx,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharasy,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Submitted (15-OCT-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 103157)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Howng,B., Kim,C.,
Koo,T., Lee,J.M., Lenx,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharasy,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Submitted (01-JUN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 103157)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.
and Davis,R.W.
Submitted (21-JUN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
5 (bases 1 to 103157)
FEDERSPIEL,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.
and Davis,R.W.
Submitted (01-SEP-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Jun 21, 2000 this sequence version replaced gi:8050909.
Bases 1-5,179 of clone T6J4 overlap with bases 94,580-99,758 of IGF
BAC clone F3F19 (gb|AC007357)
e-mail for correspondence: arab@sequence.stanford.edu Genes with
similarity to proteins in the databases are named 'putative',
'-like' or 'similar to'. Genes that have EST similarity but no
significant protein similarity are described as 'unknown proteins'.
Genes that are annotated based only on gene prediction software
are described as 'hypothetical proteins'. The gene prediction
programs used to predict genes include: Grail (Informatics Group,
Oak Ridge National Laboratory,
http://combio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
http://genes.mit.edu/GENSCAN), FEXA (Victor Solovyev,
http://genomic.sanger.ac.uk/gf/gf.shtml), and NetPlantGene (S.M.
Hebegaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).

FEATURES
Location/Qualifiers
1..103157
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T6J4"
misc_feature
1..5179
/note="overlap with bases 94,580-99,758 of IGF BAC clone
F3F19 (gb|AC007357). See GenBank record for BAC F3F19 for
annotation in this region."
complement(6893..8112)
gene

[illegible]

Db 83425 GATCGCTTAAAGTACCGTGAAGAGTCTCCTTGCAGAGTCTGCTGTCAGGAAC 83484
 Qy 568 CTAAGAGTACATTTGCTCTTTTGGAAATTTGCGAATCTGATTAACCCCGCAAGTTA 627
 Db 83485 GTGAAGACGATTTGATATCTCTTTGGTGTTCGAATCTCCATGATGGAGAAACAAT 83544
 Qy 628 TTGGACCTCGAAAGAGCGTTGATATCTTCGCATCGGGAATCTCCTTTGTGGACCGGAA 687
 Db 83545 TTAAGTCTCGAGGCTTTGTTCCGAGTCTTATGTAGCGAGTTAACTTATCGGTCCTTA 83604
 Qy 688 TGG---AATTTCCAGCGGAGTTCAACATCGCGGCTCTACTAATCAAACTTCTATTG 744
 Db 83605 TGGAAATGTTTAGGCTATTTGAGAGGAGAGCTTGGAAAGATCTTAAGCATGTGCTTTTC 83664
 Qy 745 TTGAGTTACGAAGAAATGTTAGAGAAGCCAGTTCGAAATGTGAAGAGCTAGCTGAGTTC 804
 Db 83665 TTGAGTACGAGGAGTTGAGACCGAGGCTCGTGTGCAATCAAGAGACTTGCAGAGTTC 83724
 Qy 805 ATGGATGTGGGTTTCAGACGATGAGAGAGAAACAAGGATTTGTGATGATAGTATGAA 864
 Db 83725 TTAGATTGTCATTTCAACAAGGAGAGAGATAGTGGAGGTGTAGACAAGATCTTGAA 83784
 Qy 865 CTTTGTAGCTTCGACATCTGAACAATCAACAGGTGAACAAAACGATCAAGTCAAT 924
 Db 83785 CTTTGTCTCTAAGAAACCTTAGCGGTTTGAGATCAACAAAACAGG---AAGCTTGTCG 83841
 Qy 925 TCGAAATCGCACAAACAGCATTTCTTCAGGAAAGGTGAGGTGAGAGATTCGGCAAACTAT 984
 Db 83842 GAAGGAGTAAGTTTCAAGAGTTTTCGTAAGGGAAGTGTGATGGAAGAGTTAT 83901
 Qy 985 CTAAGCTCGGAATGATTAGAAACTCGAGAGC 1017
 Db 83902 ATGACTCCTGAAATGGAAACAAAATCGCATG 83934

RESULT 7

AY085411
 LOCUS 1212 bp mRNA linear PLN 25-JUN-2002
 DEFINITION Arabidopsis thaliana clone 150484 mRNA, complete sequence.
 ACCESSION AY085411
 VERSION AY085411.1 GI:21404121
 KEYWORDS FLI CDNA.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1212)
 Haas,B.J., Volkovsky,N., Town,C.D., Troukhan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
 Full-length messenger RNA sequences greatly improve genome annotation
 Genome Biol. (2002) In press
 2 (bases 1 to 1212)
 Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.
 Full-length cDNA from Arabidopsis thaliana
 3 (bases 1 to 1212)
 Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.
 Direct Submission
 Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA
 COMMENT This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be

5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genet carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.

FEATURES

source

Location/Qualifiers

1..1212

/organism="Arabidopsis thaliana"

/db xref="taxon:3702"

/clone="150484"

31..1026

/codon_start=1

/product="steroid sulfotransferase, putative"

/protein_id="AAM62638.1"

/db xref="GI:21553545"

/translation="MGEKDIPLNKEEEREENOSSEETKSLISSLPDIDCGSGTKLY KYGCGWYKDILOAILNENKQFQSETDIIIVASPKSGTTWLKALTFAAQSKHSD NHPLLTNHPHLPVLELDLYKSKSPDULKLPSSPRLPSTHMSFDALVKPKSPC KIVYCRNVNDVLIISLWCFENSGNNLSLEAFESLCSGVNLCGLWENLVGWRG SLEDPKHVLFLEBELKTEPRVQIKRLAEFLDCPFTKEEDSGVDKILELCSLRNLS GLEINKTGSLSSEGVSKSPFRKGEVDMKSYMTPEMENKIDMIVEKLGSGGLK"

BASE COUNT 382 a 215 c 272 g 343 t

ORIGIN

Query Match 10.8%; Score 128.6; DB 8; Length 1212;
 Best Local Similarity 48.4%; Pred. No. 3.2e-19;
 Matches 452; Conservative 0; Mismatches 469; Indels 12; Gaps 3;
 Qy 88 AAGAATGACCAAGAGAGGAAGATTCCTCAAAATGTACAAGAGATATAGAGAGATTGTT 147
 Db 61 AAGGAGAAGAAGAAGAAGAAACCAAGAGTGAAGAAACCAAAAGTTGATCTCT 120
 Qy 148 TCTTCATCTCCCTCGAATGATTATTGGGGGATACCATGAGGTGTGTACAAGGATTTGG 207
 Db 121 TCACCTCTTCAGACATAGATTGCTCAGGAGCAAGTTGTACAAGTACCAAGATGTTGG 180
 Qy 208 CAAATGGGATATCTTGTTACCTGGTATCATGGCTTTCGAAAGATAATTTCAAGGCTCGAG 267
 Db 181 TAGCACAAGATATTTCTCCAAGCAATCTCTCAATTTCAACAAAACATTTTCAGCACAAGAA 240
 Qy 268 ACGGACATTTATCTTACGACTCTTCCAAAGGCTGGAAGACATGACGAGGACATGACG 327
 Db 241 ACGGATATAATTTGTTGCTCTTTCCTCCAAATCGGTAGACTTGGCTCAAGGACATCA 300
 Qy 328 TTGCGCATCTTAACACGAGATGTTTAAACCAACCCATCATCCGACACATCCATTTGTTTC 387
 Db 301 TTGCGCACTCGGCAAGA-----TCAAAACACACTTCAGACAATCATCTCTGCTAACT 354
 Qy 388 TTCACACCTCATCTCGTGTGTTCAAAATTTGGAGTATTTGTATCGGTAGAGAAATACG 447
 Db 355 CATAAATCTCTAGAGCTAGTCCGATACCTCGAGCTCGATCTTTATCTCAAAAGCTCGAA 414
 Qy 448 ATGCCAGACCTCGATATGTTGAATGAATCGCGAGGTTGTTGCGGACACATCCCATAC 507
 Db 415 CGGATTTGACCAAGTTCGCCATCATCTCGAGATTTGTTCTCAACCCACATGTCCTTT 474
 Qy 508 TCTTTGTTGCGGCGTCTGTTTGAATTCGGGAAACAAAATCATCAATATATAGCGCGAAC 567
 Db 475 GATGCGCTTAAAGTACCGTTCGAAAGAGTCTCTCTTCAAGATCGTGTACGTGTGCGAGGAA 534
 Qy 568 CGTAAGAGTACATTTGCTCTTTTGGAAATTTGGCATCTGATTAACCCCGACAGTTA 627
 Db 535 GTGAATGACGATTGATATCATCTTTGGTGTTCGAAATCTCCATGAGTGGAGAAACAAT 594
 Qy 628 TTGACCTCGAAAAGAGCGTTGATATCTTCGCATCGGGAATCTCTCTTTTGGAGCGGAA 687
 Db 595 TTAAGTCTCGAGGCTTTGTTGAGTCTTTATGTAGCGAGTAACTATATGCGGTCCCTTA 654
 Qy 688 TGG---AATTTCCAGCGGAGTTTCAACCAATCGGCGTCTACTAATTCAAACTTCTATTG 744

Db	655	TCGGAAATGCTGTTAGGCTATTGGAGGAGGAGCTTGAAGATCCTAAGCATGTGCTTTTC	714	Qy	161	CGAATGATTATTGGGGGATACCATGATGGTGTCAAGGGATTTTGGCAATGGGATATC	220
Qy	745	TTGAGTTACAGAAATGTTAGAGAGCCAGTTGAAATGTGAAGAGTACGTGATTC	804	Db	98	ACAAAGATTTCATGGTTATGGTCTCTACAACTACAAAGTTGTGGTACTATCAACA	157
Db	715	TTGAGGTACGAGGAGTTGAAGACGAGCCCTCGTGTGCAAAATCAAGAGACTTGCAGATTC	774	Qy	221	TTGTACCTGGTATCATGGCTTTCGAGNATAATTTCAAGGCTCGAGAGACGACATATCC	280
Qy	805	ATGGGATGGGTTCACAGACGATGAGGAGAAACAAGGGATTGTTGATGAGATAGTTAA	864	Db	158	CATCTCAAGCCGTTCTTGACGTCCAAAACACTTCAAGCCAGAGATACTGATATAATCC	217
Db	775	TTAGATTGCCATTCAAGAGGAGAGAGATAGTGGAGGTAGACAAGATCTTGGAA	834	Qy	281	TTACGACTCTTCAAGGCTGGAACGACATGGAGGAAGGCACTGACGTTTGGCCATCTAA	340
Qy	865	CTTTGTAGCTTCGACAAATCTGAAGATCAACAGGTCAACAAAACGGATCAAGCTACAAT	924	Db	218	TCGCTTCTTGGCCAAAGGTGGACCACTTGGCTCAATCCTTAATTTTCGCTGTGTAC	277
Db	835	CTTTGTCTTAAGAAACCTTAGCGTTTGGAGATCAACAAAACAGG---AAGCTTGTGC	891	Qy	341	CACGAGATGTTAAACCAACCCATCATCCGACACATCCACTTTTGTTCCTTCAACCTCAT	400
Qy	925	TCGAAATCGACAAACAGCATTTCTTCAGGAAGGTGAGTGAGAGATTGGGCAACTAT	984	Db	278	ATAGAGAAAGATGACCGCGAACCCCTCAACACATCTTTTGTCTTACAAAACCCCTCAT	337
Db	892	GAAGGAGTAGTTTCAAGAGTTTTCCTCGTAAAGGGGAAAGTTGGTATTGGAGAGTTAT	951	Qy	401	CGTGTGTTCAAAATTTGGAGTATTTGTACATGGGTAGAGAAATACGATGCCAGACCTCG	460
Qy	985	CTAACGTCGGAATGATTAAGAACTGGAGACG	1017	Db	338	ACCTTGTCCCATTTCTTGAGGTTGATATACGCTA---ATAGCCAAATTCGGATCTCG	394
Db	952	ATGACTCTGAAATGGAAACAAATTCGACATG	984	Qy	461	ATATGTTGAATGAATCGCGGAGGTTTGTTCGCGGACACATCCCATCTCTTTTGTTCGGG	520
RESULT 8	AF325065	996 bp mRNA linear	PLN 11-MAR-2001	Db	395	CAAAGT---ATTCTTCTCTATGATCTTTCTACACACATGCACCTTACAAGCATTTGCGTG	451
LOCUS	Arabidopsis thaliana putative steroid sulfotransferase (At2g03750)			Qy	521	CGTCTGTTTGAATCGGGAACAAAATCATCAATATAAGCCGCAACCGTAAGATGATACAT	580
DEFINITION	mRNA, complete cds.			Db	452	AAGCCACCACAAAAGCTTG---CAAACCGTATATGTGTGTAGAGGTATCAAAAGATACGT	508
ACCESSION	AF325065			Qy	581	TTGTGCTTTTTGGAAATTTGGCAATCTGATTAAACCCGACAAAGTT-----ATTGGACC	634
VERSION	AF325065.1	GI:13272388		Db	509	TTGTCTCCGCTGGCATTTATAGAAACATGTTCCATCGCAACCAAGATGGATCAAGCCACTT	568
KEYWORDS	FLI CDNA.			Qy	635	TCGAAAGAGCGTGTGATATCTTCGCATCGGGAATCTCTTTTGTGGACCGGAATGGAAT	694
SOURCE	Arabidopsis thaliana.			Db	569	TTGAGCTCATGTTTGTATGCTTATTTGTAGAGGAGTTCTCTTATATGACCTTATTTGGAAC	628
ORGANISM	Arabidopsis thaliana.			Qy	695	TCCAAGCGGAGTTCAACCAATGCGGGCTCT---ACTAATTCAAACTGCTATTGTTGAGTT	751
REFERENCE	1 (bases 1 to 996)			Db	629	ATGTTATGAGCTATTGGAAAGGAGCTTGGAAAGGAGGATGTTCTTTTCATGAAGT	688
AUTHORS	Southwick, A., Karlin-Neumann, G., Nguyen, M., Lam B., Miranda, M., Palm, C.J., Theologis, A., Ecker, J. and Davis, R.W.			Qy	752	ACGAGAAATGTTAGAGAGCCAGTTGAAATGTGAAGAGCTAGCTGAGTTTCATGGGAT	811
TITLE	Direct Submission			Db	689	ACGAGAGATAATTGAGGAGCTCTGTTCAAGTCAAGAGACTCGCGGAGTTCTTGAAT	748
JOURNAL	Submitted (30-NOV-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA			Qy	812	GTGGGTTTCACAGCCGATGAGGAGAAACAAGGATGTTTTCATGAGATAGTTAACTTTGTA	871
FEATURES	Location/Qualifiers			Db	749	GTCCATTACCAAGGAAGAAGAAAGTGGATCGGTGGAGGAGATCTTGAAGTTGTGTA	808
source	1..996	/organism="Arabidopsis thaliana"		Qy	872	GCTTCACAACTGGAAGATCAACAGGTGACAAACCAACGATCAAGCTACAATTCGAAA	931
gene	1..996	/db_xref="taxon:3702"		Db	809	GTTTACGNAATTTAAGCAATTTGGAGGTTAATAAGAAATGGGACACGAGAAAT---GGTG	865
CDS	1..996	/clones="c6004"		Qy	932	TCGACAAACAGCATTTCTTCAGGAAAGGTGAGGTGAGAGATTGGGCAACTATCTAACGT	991
	1..996	/gene="At2g03750"		Db	866	TAGATTCTCAGGTGTTCTTTAGGAAAGGTGAGTTGTTGTTGGAAGAATCATCTTACGC	925
	1..996	/gene="At2g03750"		Qy	992	CGGAAATGATTAAGAAACTGGA	1013
	1..996	/codon_start=1		Db	926	CACAATGGCGAAACCTTTGA	947
	1..996	/product="putative steroid sulfotransferase"		RESULT 9			
	1..996	/protein_id="AAK17133.1"		AY087921			
	1..996	/db_xref="GI:13272388"		LOCUS			
	1..996	/translation="MEASKEAHLPNWKDNVSOETKNLITSLPSDKDFWGYGLNY		Arabidopsis thaliana clone 3964 mRNA, complete sequence.			
	1..996	KGCHYNTLQAVLDVQKHKPRDITLILASLPKGGTTWLKSLIFAVVHREKYRGTPQ		DEFINITION			
	1..996	THPLLQNPDLVPFLEVLYANSQIEDLAKYSSPMIFSTHMLQALREATTACACTV		Arabidopsis thaliana			
	1..996	YVCGIKDTFVSGWHYHMLHRTKMDQATFELMFDAYCRGVLLYGPTEHVLSTWKS		AY087921			
	1..996	LEAKENLPMKYBIIIEPRVQVKLAEFLCEPTEEBESGSVEEILKLCSLNLSN		Accession			
	1..996	LEVKNKGTTTRIGDSQVFFRKGEVGDWKNHLTPQAKTFDBIIDYRLGSLIFQ"		AY087921.1			
BASE COUNT	306 a 199 c 213 g 278 t			VERSION			
ORIGIN				GI:21406695			
Query Match	10.7%	Score 127.2;	DB 8;	KEYWORDS			
Best Local Similarity	50.8%	Pred. No. 6.9e-19;		FLI CDNA.			
Matches 468;	Conservative	0;	Mismatches 433;	SOURCE			
			Indels 21;	ORGANISM			
Qy	101	AGAGGAGAGATTCCAAATGTACAAGAGATATAGAGAGATTGTTTCTTCACTTCCT	160	Arabidopsis thaliana			
Db	38	ACATGAAAGACGACAACTGTTAGTCAAGAAACCAAGAACTTATCTTCTACCTTCAG	97	Embryophyta; Tracheophyta;			
				Eukaryota; Viridiplantae;			
				Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
				Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 1154)
Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N.,
Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
Genome Biol. (2002) In press
2 (bases 1 to 1154)
Brower,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 1154)
Brower,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the Ws or Laer ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Genet carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.
Location/Qualifiers
1. 1154
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="3964"
93.11088
/codon_start=1
/product="putative steroid sulfotransferase"
/protein_id="AAM65471.1"
/db_xref="GI:21593504"
/translations="MEASKEAHLPMYKDDNVSOETKNLITLPSDKDFMGVGLVNY
KGVWYNTLQAVLDVORHFKPRDITLILASLPKGGTTLKSLIFAVHREKVRTPQ
THELLQNPHDLVFEVLEYANSQIPDLAKYSPMIFSTHMLQALREATTACKATV
YVCRGIDTFSVGMHVMHLRMTQDQATPELMFDAYCRGVLLGYVWEHVLVSYKGS
LEAKENLVFKYBEIIEEPVQVVKRLAEFLCEPFTKEEESGVSVEILKLSRLNLSN
LEVNKGTTTRIGVDSOVFFPRKGVGMKNHLTPQMAKTFDEIIDYRLGDSGLIFQ"

BASE COUNT 353 a 226 c 235 g 340 t
ORIGIN

Query Match 10.7%; Score 127.2; DB 8; Length 1154;
Best local Similarity 50.8%; Pred. No. 6.8e-19;
Matches 468; Conservative 0; Mismatches 433; Indels 21; Gaps 6;

QY 101 AGAAGGAAGAGATTCCAAATGTACAGAGATATAGAGAGATTGTTCTTCACTTCCT 160
DB 130 ACATGAAGACGACCAAGCTTAGTCAAGAAACCAAGAACTTGATCACTTCTACCTTCAG 189
QY 161 CGAATGATTATGGGGGATACCATGAGGTGTGTACAAAGGATTTTGGCAATGGGATATC 220
DB 190 ACAAGATTTCATGGGTATGTTCTCTACACTCAAGAGTTTGTGGTACTATCCAAACA 249
QY 221 TTGTACTGTATCATGGCTTTCCAGAGATAATTTCAAGGCTCGAGAGCGGACATTATCC 280
DB 250 CACTCAAGCGCTTCTTGAGCTCCAAAACACTTCAAGCCACGAGATACGTATATATCC 309
QY 281 TTACGACTCTTCCAAAGCTGGGAACACATGAGCAAGGACGTGAGTTTGGCATCTTAA 340
DB 310 TCGCTTCTTTGCCCAAGGTGGAACCACTTGGCTCAAAATCCCTAATTTTCGCTGTGTAC 369
QY 341 CACGAGATGTTAAACCCCATCATCACCGACACATCCACTCTTTTGTCTTCAACCCCTCAT 400

Db 370 ATAGAGAAAGTACCGCGGAACCCCTCAAAACATCTTGTCTTACAAAACCCCTCATG 429
QY 401 CGTGTGTTCAAAATTTGGAGTATTTGTACATGGGTAGAGAAAATACGATGCCACAGACTCG 460
DB 430 ACCTTGTCCTCATTTCTTGAGGTTGAGTTATACGCTA--ATAGCCAAATTTCCGATCTCG 486
QY 461 ATATGTGTATGAATCGCGGAGTTTGTTCGCGGACATATCCCATACTCTTTTTCGCGG 520
DB 487 CAAAGT---ATTCTTCTCTATGATCTTTCTACACATATGCACTTACAGCAATTCGCTG 543
QY 521 CGTCTGTTTGAATCGGGAACAAAATCATCAATATAAGCCGCAACCGTAAGAGTACAT 580
DB 544 AACCCACCAAAAGCTTG---CAAAACGATATGTGTGTAGAGGTATCAAGATACGT 600
QY 581 TTGTGTTCTTTTGGAAATTTGGCAATCTGATTAACCCGCAAGTT-----ATTGACC 634
DB 601 TTGCTCTCGGCTGGCAATTATAGAAACATGTGTCATCGCAACAGATGGATCAAGCCACTT 660
QY 635 TCGAAAAGAGCGTTGATATCTTCGCATCGGGAATCTCTTTTGTGACCGGAATGGAAT 694
DB 661 TTGAGCTCATTTTGAAGCTTATTTGAGGAGTCTCTTATATGAGACCTTATTGGGAAC 720
QY 695 TCAACGCGAGTTTCACAATGCGGCTCT---ACTAATTTCAAACTTGTCTTATTGTTGAGTT 751
DB 721 ATGTATTGAGTATTGGAAGGAGCTTGGAAAGCAAGGAGATGTTCTTTTCATCAAGT 780
QY 752 ACGAAGAAATTTAGAGAACCGAGTTGAAAATGTGAAGAGCTAGCTGAGTTTCATGGGAT 811
DB 781 ACGAAGAGATAATTTGAGGAGCTCGTGTTCAGAGTCAAGAGACTCGCGGAGTTCTTGGAA 840
QY 812 GTGGGTTACACAGCATGAGGAGAAACAAGGATTTGTATGATGAGATAGTTAACTTTGTA 871
DB 841 GTCCATTACCAAGGAAGAAAGAGTGGATCGGTGGAGGAGATCTTGAAGTTGTGTGA 900
QY 872 GCTTCGCAATCTGAAGATCAACAGGTGAAACAAAACCGATCAAGTCAACAATTCGAAA 931
DB 901 GTTTACGAAATTTAAGCAATTTGGAGTTTAAAGATGGGACACAGAGAAT---GGTG 957
QY 932 TCGACACACAGCATTTCTTCAGAAAGGTGAGGTGAGAGATTGGGCAACTATCTAAAGT 991
DB 958 TAGATTCTCAGGTGTTCTTTAGGAAAGGTGAAGTTGGTGAATGGAAGATCATCTTACGC 1017
QY 992 CGGAAATGATTAAAGAACTGGA 1013
DB 1018 CACAATGCGAAACCTTTGA 1039

RESULT 10
AY092961
LOCUS
DEFINITION
Arabidopsis thaliana putative steroid sulfotransferase (At2G03750)
mRNA, complete cds.
AY092961
AY092961.1 GI:20260123
VERSION
FLI CDNA.
KEYWORDS
SOURCE
Arabidopsis thaliana
Arabidopsis thaliana
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1216)
Southwick,A., Karlin-Neumann,G., Nguyen,M., Lam,B., Miranda,M.,
Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (27-MAR-2002) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
e-mail for correspondence: arab@sequence.stanford.edu

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: "RIKEN Arabidopsis Full-length cDNA"): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.

The Salk, Stanford, PDEC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Southwick, A., Nguyen, M., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C. J., Bower, L., Jones, T., Ban, J., Chen, H., Cheuk, R., Chung, M. K., Kim, C., Lin, J., Liu, S. X., Pham, P. K., Sakano, H., Shinn, P., Yamada, K., Ecker, J., Theologis, A., and Davis, R. W.

Southwick, A., (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R. W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES		Location/Qualifiers	
source	1..1216	/organism="Arabidopsis thaliana"	
		/db_xref="taxon:3702"	
gene		/chromosome="2"	
		/clones="RAPL04-17-K21 (R15664)"	
CDS		/notes="This clone is in pBluescript"	
		ecotype: Columbia"	
	1..1216	/gene="At2g03750"	
		/notes="F19B11.20"	
	42..1097	/codon_start=1	
		/product="putative steroid sulfotransferase"	
		/protein_id="AA12960.1"	
		/db_xref="GI:20260124"	
		/translation="MFTFTILSLCFKSWQIITMEASKAHLPLNPKMDNVISQETK NLITSLSDKDFMGYKLYNQKGYNTLQAVLDVQKFKPRDIDIIILASLPKGGTT WLKSLIFAVVHREKRYRTPOTPLLLQHPDLVPLEVLYANQIPDLAKYSMPIF STHMLQALREAVTKACKTVVCRGIKDITFVSGWYHRLMHTKMDQATPELMFDAYC RGVLLYGPYWEHLVSKGSLKAKENLVLPKVEIIEPRVOKRLAEFLCEPFTKEE EESGSVEILKLSRLNLSLVNKNKGTTRIGVDSQVFFRKGEGVDKWNHLLTPQMAKT PDEIIDLKSDGSLIFQ"	
		BASE COUNT 365 a 234 c 245 g 372 t	
ORIGIN			
Query Match		10.7%; Score 127.2; DB 8; Length 1216;	
Best Local Similarity		50.8%; Pred. No. 6.8e-19;	
Matches 468; Conservative		0; Mismatches 433; Indels 21; Gaps 6;	
QY	101	AGAAGGAAGAAGATTCCAAAATGTACAAAGATATAGAGAGATTGTTTCTTCACTTCCCT	160
DB	139	ACATGAAGACGACACACCTTAGTCAAGAAACCAAGACTTGCATCTCTCTACCTTCAG	198
QY	161	CGAATGATTATTTGGGGGATACCATGAGTTGTACAGGGATTTTGGCAATGGATATC	220
DB	199	ACAAAGATTTCATGGTTTATGCTCTCTACAACTACAAAGGTTTGTGTACTATATCCAA	258
QY	221	TGTACCTGTTATCATGCTTTCGAAGATAATTTCAAGGCTCGAGACGACGACATTATCC	280
DB	259	CACCTCAAGCCGCTTCTTACGCTCCAAAACACTTCAAGCCAGGATGATGATATATCC	318
QY	281	TTACGACTCTTCCAAAGGCTGGAACGACATGGACGAAGGCACTGACGTTTCCCATCTAA	340
DB	319	TCGCTTCTTTGCGCAAGGTGGAACCACTTGGCTCAAACTCCATAATTTGCTGTGTAC	378
QY	341	CACGAGATGTTAAACCAACCATCATCAGCAGACATCTTCTTCTTCAACCTCATTT	400
DB	379	ATAGAGAAAAGTACCGCGGAACCCCTCAACACATCTTCTTCTTCAAAACCCCTCATG	438
QY	401	CGTGTGTTCAAAATTTGGAGTATTTGTACATGGGTAGAGAAATACGATGCCAGACCTCG	460
DB	439	ACCTTGTCCCATTTCTTGAGTTGAGTTATAGCTA---ATAGCCAAATTCGGATCTCG	495
QY	461	ATATGTTGAATGAATCGCGAGGTTGTTTGGCCGGACACATCCCATCTCTTTTGTGGCGG	520

DB	496	CAAACT---ATTCTTCTCTATGATCTTTTACACACATGCACCTTAAAGCATTTGGCTG	552
QY	521	CGTCTGTTTGAATTCGGNAACAAAATCATCAATATAAGCCGCAACCGTAAGAGTACAT	580
DB	553	AAGCCACCACAAAAGCTTG---CAAAACCGTATATGTGTGTAGAGGTATCAAGATACGT	609
QY	581	TTGTCTCTTTTGGAAATTTGGCAATCTGATTAAACCCCGACAAGTT-----ATTGGACC	634
DB	610	TTGTCTCCGGCTGGCATTTATAGAAACATGTTTGCATCGCACCAAGATGGATCAAGCCACTT	669
QY	635	TGAAAAGAGCGTTCATATCTTCGCATCGGAATCTCCTTTTGTGGACCGGAATGGAATT	694
DB	670	TTGAGCTCATGTTTGTATGCTTTATTGTAGAGGAGTTCTCTTATATGGACCTTATTGGGAC	729
QY	695	TCCAAGCGGAGTTCCCAATCGCGCTCT---ACTAATTTCAAACTTGTCTTATTGTTGAGTT	751
DB	730	ATGTTATTGAGCTATTGGAAAGGAGCTTGGAAAGCAAGAGGAGATGTTCTTTTCATGAAGT	789
QY	752	ACGAAGAAATTTAGAGAACGCCAGTTGAAAATGTGAAGAGCTAGCTGAGTTTCATGGGAT	811
DB	790	ACGAAGAGATATTGAGGAGCCTCGTGTTCAGTCAAGAGACTCGCCGAGTTCTTGGAAAT	849
QY	812	GTGGGTTACACACGATGAGGAGAAACAGGAGTTGTTGATGAGATAGTTAACTTTGTA	871
DB	850	GTCCATTACCAAGGAAGAAGAGTGGATCGGTGGAGAGATCTTTGAAGTTGTGTA	909
QY	872	GCTTCGACAACTCTGAAGATCAACAGGTGAACAAAACGGATCAAGCTACAAATTCGAAAA	931
DB	910	GTTTACGAATTTTAAGCAATTTGGAGGTTTAAAGATGGGACACAGAGAAT---GGTG	966
QY	932	TCGACAAACAGCATTTCTTACGAAAGGTGAGGTGAGAGATTGGGCAAACTATCTAACT	991
DB	967	TAGATTCTCAGGTGTTCTTTAGGAAAGGTGAAGTTGGTGAATGGAAGATCATCTTACGC	1026
QY	992	CGGAATGATTAAAGACTGGA 1013	
DB	1027	CACAAATGCGAAAAACCTTTGA 1048	

RESULT 11
AC006836/c Arabidopsis thaliana chromosome 2 clone F19B11 map RNSI, complete
LOCUS AC006836
DEFINITION AC006836
ACCESSION AC006836
VERSION AC006836.7 GI:20197898
KEYWORDS HTG
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 87543)
Lin, X., Kaul, S., Shea, T. P., Fujii, C. Y., Shen, M., VanAken, S. E.,
Barnstead, M. E., Mason, T. M., Bowman, C. L., Renning, C. M., Town, C. D.,
Benito, M. I., Carrera, A. J., Creasy, T. H., Buell, C. R., Town, C. D.,
Nierman, W. C., Fraser, C. M. and Venter, J. C.
Unpublished
2 (bases 1 to 87543)
Lin, X.

REFERENCE 3 (bases 1 to 87543)
Town, C. D. and Kaul, S.
Direct Submission
Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdownt@igir.org

REFERENCE 4 (bases 1 to 87543)
Town, C. D. and Kaul, S.
Direct Submission
Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdownt@igir.org
On Apr 18, 2002 this sequence version replaced gi:6598617.
On Oct 8, 1997 this sequence version replaced gi:2443866.
We have determined that YAC YUP8H12 is chimeric, and is comprised
of two distinct genomic EcoRI fragments from chromosome 1. This

mRNA
 join(16077..16503,16583..16721,16810..16963,17036..17159,17250..17624,17706..18065)
 /genes="At2g03590"
 CDS
 join(16276..16503,16583..16721,16810..16963,17036..17159,17250..17624,17706..17858)
 /genes="At2g03590"
 /codon_start=1
 /product="expressed protein"
 /protein_id="A02067.1"
 /db_xref="GI:4406756"
 /translation="MYMIESKGAIAICMLALLFLGTWPAIMTLTERRGLRPOHYLDYTNLNLAAVITAIATLTGEGPSRPNFEQTQSNDQMSVFMFAMAGIVLSLGNATQYA WYVGLSVTEVITATITVIGTTLNFDLRINRAEVLFPVACFLIAVFCGSAIVHKS NAADNTKLNQKFSLETTSSPEMETISASNGLTGKAKEGTAFLIELEKORAIKVFG KSTIIIGLVITTFAGICFSLFSPFNLATNDQWHTLKHPKPLNVTAFYFISIAFV ALINTFLFYPLIGLPRFSKAYLNDWNGRWSFLAGFLCGFNGLQFMGQAGAYA AADAVQALPLVSTFWGILLFGEYRRSRKTYTLISMLLMFIVAVAVLMASSGHRK"
 18441..19910
 /genes="At2g03600"
 /note="F19B11.5"
 join(<18441..18668,18738..18876,18953..19106,19188..19299,19377..19425,19524..19774,19809..19910)
 /genes="At2g03600"
 Query Match
 Best Local Similarity 50.8; Pred. No. 5.8e-19;
 Matches 468; Conservative 0; Mismatches 433; Indels 21; Gaps 6;
 10.7%; Score 127.2; DB 8; Length 87543;
 70073 ACATGAAAGACGCAACGTTAGTCAAGAAACCAAGAACTTGTATCTCTACCTTCAG 70014
 161 CGAATGATTATGGGGGATACCATGAGTTGTACAGGATTTTGGCAATGGATATC 220
 70013 ACAAGATTTTCATGGGTTATGGTCTCTACAACTACAAGGTTGTGTATCTATCCAAACA 69954
 221 TTGTACCTGGTATCATCGCTTTTCAAGATAATTTCAAGGCTCGAGACGCGACATTATCC 280
 69953 CACTCCAAGCGTTCTTGACGCTCCAAAACACTTCAGCCACGAGATCTGATATATCC 69894
 281 TTACGACTCTTTCCAAAGCTGGAAACGACATGAGCAAGGCACTGACCTTTGCCATCCTAA 340
 69893 TCGCTCTTTGCCAAAGGTGGAACCACTTGGCTCAAAATCCCTAAATTTTCGCTGTGTAC 69834
 341 CACGAGATGTTTAAACCCATCATCCGACACATCTTCTTCTCAACCTCAT 400
 69833 ATAGAGAAAGTACCGGGAACCCCTCAACACATCTTCTCTTCAAAACCTCATG 69774
 401 CGTGTGTTTCAAAATTTGGAGTATTTGTATCGGTGAGAGAAATACGATGCCAGACCTCG 460
 69773 ACCTTGTCCTATTTCTTGAGGTTGAGTTATACGCTA---ATAGCCAAATTCGGATCTCG 69717
 461 ATATGTTGAATGAATCGCCGAGGTTGTTGGCGGACACATCCCATCTCTTTGTCGGG 520
 69716 CAAAGT---ATTCTCTCTATGATCTTTTCTACACACATGACCTTCAAGACATTGCGTG 69660
 521 CGTCTGTTTGAATCGGGAACAAAATCATCAATATAAGCGCAACCGTAAGAGTACAT 580
 69659 AAGCACCACAAAAGCTTG---CAAAACCGTATATGTTGTAGAGGTATCAAGATACGT 69603
 581 TTGTGCTCTTTTGGAAATTTGGCAATCTGATTAACCCCGACAAGTT-----ATTGGACC 634
 69602 TTGCTCCGGCTGGCATTATAGAAACATGTTGTCATCGCACCAAGATGATCAAGCACTT 69543
 635 TCGAAAAGACGCTGATATCTTCGACATCGGNAATCTCTTTTGGACCGGAATGGAAATT 694
 69542 TTGAGCTCATGTTTGATGCTTATTGTAGAGGAGTTCTCTTATATGACCTTATTGGGAAC 69483
 695 TCCAAGCGAGTTTACCATCGGCGTCT---ACTAATTTCAAACCTTGCTATTGTTGAGTT 751
 69482 ATGTTATGAGCTATTGGAAAGGAGCTTGGAGCAAGAGGAAATGCTTCTTTTCATGAGT 69423
 752 ACGAAGAAATGTTAGAGAAGCCAGTTGAAATGTGAAGAGCTAGCTGAGTTTCATGGAT 811

Db
 69422 ACGAAGAGATAATTGAGGAGCCTCGTGTTCAGTCAAGAGACTGCCGAGTTCTTGGAA 69363
 Qy
 812 GTGGGTTACAGAGCATGAGGAGAAACAAGGATTTGTTGATGAGATAGTTAAACTTTGTA 871
 Db
 69362 GTCCATTTCCACAGGAAGAAGAAGTGGATCGGTGGAGGATCTTGAAGTTGTGTA 69303
 Qy
 872 GCTTCGACAATCTGAAGAATCAACAGGTGAACAAAAACGGATCAAGCTACAAATTCGAAAA 931
 Db
 69302 GTTTACGAAATTTAAGCAATTTGAGGTTTAATAAGATGGCAACGAGAATT---GGTG 69246
 Qy
 932 TCGACAAACAGCATTTCTTACGAAAGTGTAGGTGAGAGATGGCCAACTATCTAACGT 991
 Db
 69245 TAGATTCTCAGGTGTTCTTTAGGAAAGTGAAGTTGGTGATTGGAAGAAATCATCTTACG 69186
 Qy
 992 CGGAATGATTAAGAAACTGGA 1013
 Db
 69185 CACAAATGCGGAAAAACCTTTGA 69164
 RESULT 12
 AC006232 90341 bp DNA linear PLN 11-MAR-2002
 LOCUS Arabidopsis thaliana chromosome 2 clone F10A12 map B68, complete
 DEFINITION sequence.
 AC006232 AC006232.4 GI:20197647
 VERSION AC006232
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1 (bases 1 to 90341)
 Lin.X., Kaul.S., Shea.T.P., Fujii.C.V., Shen.M., VanAken.S.E., Bartsch.M.E., Mason.T.M., Bowman.C.L., Ronning.C.M., Benito.M.-I., Carrera.A.J., Creasy.T.H., Buell.C.R., Town.C.D., Nierman.W.C., Fraser.C.M. and Venter.J.C.
 Unpublished
 2 (bases 1 to 90341)
 Lin.X.
 JOURNAL Direct Submission
 REFERENCE Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 AUTHORS Medical Center Dr., Rockville, MD 20850, USA
 JOURNAL
 3 (bases 1 to 90341)
 Town.C.D. and Kaul.S.
 REFERENCE Direct Submission
 27-PER-2002 The Institute for Genomic Research, 9712
 JOURNAL Medical Center Dr., Rockville, MD 20850, USA, cdton@tigr.org
 COMMENT On Apr 18, 2002 this sequence version replaced gi:6598538.
 FEATURES
 Location/Qualifiers
 1..90341
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="B68"
 /clone="F10A12"
 complement(1..31422)
 /notes="overlap with BAC clone F12K2 (AC006233.1..31423).".
 complement(31..1042)
 /genes="At2g27330"
 Ceres:18569
 complement(join(<31..443,747..853,1020..1042))
 /genes="At2g27330"
 complement(join(348..443,747..842))
 /genes="At2g27330"
 /codon_start=1
 /product="putative RNA-binding glycine rich protein (RGP-2)"
 /protein_id="AAM15185.1"
 /db_xref="GI:20197660"


```
/translation="MDKIRCRPKGFAYVYTFSSKEEAEKALLELNAQLVDGRVWILDTT
KAAKHNPDPTRKPNHAGGDS"
1312..1502
/notes="molecular marker ngall126"
1449..1481
/rpt_family="GA)n"
2447..4118
/gene="At2g27340"
/notes="F10A12.2"
join(<2447..2556,2782..2852,3021..3148,3473..3530,
3709..3925,4031..4118)
/gene="At2g27340"
join(2447..2556,2782..2852,3021..3148,3473..3530,
3709..3925,4031..4118)
/gene="At2g27340"
/codon_start=1
/product="similar to PIG-L"
/protein_id="AAM15175.1"
/db_xref="GI:20197648"
/translation="MNVVFLSLIVIVWVSFFKIFFRATISIRATLDDGKFFSPIN
YFTSTACNLHILCFSTGNADGMSIRDOELHRACAVLKVPFQKEGICDNDSCNNEE
HIITFDNYGVGHCHNRDHPDIDCKDSAKRIHGLVYVHOVSLNIPKCYGGBDILWL
SILSAKRHPKSVIIINKQPKWKSFKAMAQHLISQWVFKLFLVSSYTYVNTLDRINPE
SNEL"
10477..10509
/rpt_family="AT_rich"
complement(10627..10647)
/rpt_family="AT_rich"
11230..12902
/gene="At2g27360"
/notes="F10A12.7"
join(<11230..11488,11709..11927,12026..12174,12269..12545,
12622..12902)
/gene="At2g27360"
join(11230..11488,11709..11927,12026..12174,12269..12545,
12622..12902)
/gene="At2g27360"
/codon_start=1
/product="putative lipase"
/protein_id="AAM15186.1"
/db_xref="GI:20197661"
/translations="MASQDCHMLLSPFISFTLITVTSQTRCNFKSIISFGDSITDT
GNLLGSSPNLDESAPPYGETFFHPGSRGDLIIDFAEFIGIPHPVPPFYGSK
NGNFKGVNFAVGATALECSVLEEKHCQSQNSILGNQLKSFRESLYLCSGSSPD
CRDMENAFLLICEIGNDYNFPLFRKNIIEVEKELVPLVITISSAIELVDMGART
FLVPGNPLFCGSVAYLTLYETPNKEEYNPLTGCLTWLNDPFSVYHNSQLOAELKRLNL
YPHWIIYGYVYNTLRLMOERSKFGLMORPLPACCGLGPGYNFTSIKGSKGVEYC
SDPSKYVWDGIHMTAEAYKWISEGVLTPGYAIPPNWNSCLDSKIKNNSLSLITQYSLM
NS"
complement(13010..13032)
/rpt_family="(CAATA)n"
complement(13282..13310)
/rpt_family="AT_rich"
complement(13447..14902)
/gene="At2g27370"
/notes="F10A12.5"
complement(join(<13447..13843,14110..14233,14580..>14902))
/gene="At2g27370"
complement(join(13625..13843,14110..14233,14580..14902))
/gene="At2g27370"
/codon_start=1
/product="unknown protein"
/protein_id="AAM15182.1"
/db_xref="GI:20197657"
/translations="MDIEKAGSRREEEPIVORPKLDKKGKAHVAPAPPMVYNRIMDK
HKOEKMSPAWGKRGVAFDFVFLRLIAITAMAAAKMATETLFFTOFLQADYT
DLPTMSFSFVINSVGGYLTLSLPSFISVILRLPLAVPPRFILICDTVMNGUTLMAAS
ASAAIYLAHNGNSNWLFPVCCQFGDQCGTSGAVVASFIATLMLFLVLSAFALK
RIT"
complement(14066..14104)
/rpt_family="AT_rich"
15671..16321
/gene="At2g27375"
/translation="MEVFFCQEVVETREHLFFSRPYSTQVMSALTQDMLGLWFTTRW
ESLIPLIMASTTPOVSQASHLGIKLLASDLQQLVKALNGAPHMKLRGILHLIHLI
SLTFARISFGFVKRKNLXEDALSKAALNMISSGPI"
18408..20693
/gene="At2g27380"
/notes="F10A12.4; predicted by genefinder and genscan"
<18408..>20693
/gene="At2g27380"
18408..20693
/gene="At2g27380"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM15181.1"
/db_xref="GI:20197656"
/translation="MRVPLIDELRLVLILSLSGASVAADATVKQNFNKYETDSGAH
PPIYGAAPSYYTPPPPIYSPPIYPPPIQKPTYSPPPIYPPPIQKPTYSPPPIY
PIQKPTYSPPPIYPPPIQKPTYSPPPIYPPPIQKPTYSPPPIYPPPIYPPPIY
PYSPPIKPPPVHKPTPTYSPIKPPPVHKPTPTYSPIKPPPVHKPTPTYSPIK
PPPVHKPTPTYSPPVHKPTPTYSPIKPPPVHKPTPTYSPIKPPPVHKPTPTYS
PIKPPPVHKPTPTYSPPVHKPTPTYSPPVHKPTPTYSPIKPPPVHKPTPTYS
KEPTYSPIKPPLOKPTPTYSPIKPPPVHKPTPTYSPPVHKPTPTYSPIKPP
PVKPPPVHKPTPTYSPIKPPPVHKPTPTYSPPVHKPTPTYSPIKPPPVHKPT
OKPTYSPIKPPPVHKPTPTYSPIKPPPVHKPTPTYSPIKPPPVHKPTPTYS
SPPIKPPPVHKPTPTYSPIKPPPVHKPTPTYSPIKPPPVHKPTPTYSPIKPP
PVKPTYSPIKPPPVHKPTPTYSPIKPPPVHKPTPTYSPIKPPPVHKPTPTYS
PYSPVHKPTPTYSPIKPPPVHKPTPTYSPIKPPPVHKPTPTYSPIKPPPV
OGGYTTPPYAYLSHPIDIRN"
23038..23071
/rpt_family="AT_rich"
complement(23163..23184)
/rpt_family="AT_rich"
complement(25481..25885)
/gene="At2g27390"
/notes="F10A12.3; predicted by genefinder"
complement(<25481..>25885)
/gene="At2g27390"
complement(25481..25885)
/gene="At2g27390"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM15178.1"
/db_xref="GI:20197652"
/translation="MKMNTKEYSIRTPSLGNLPGTTGTCCPPPLVFLPLSPPPS
PPSPSPPLPPPPPLFPPEPLPPPLFPPLFPPLFPPLFPPLFPPLFPPLFPPE
PPPPPPPEPPPPPPASCLRTKSPENGIVSTVK"
25556..25663
/rpt_family="(GGA)n"
25713..25783
/rpt_family="(TGG)n"
complement(26048..26321)
/gene="At2g27395"
/notes="F10A12.8"
complement(join(<26048..26156,26170..>26321))
/gene="At2g27395"
complement(join(26048..26156,26170..26321))
/gene="At2g27395"
/codon_start=1
/product="putative cysteine proteinase"
/protein_id="AAM15183.1"
/db_xref="GI:20197658"
/translation="MNNEEALVQAVSQHPVSMIEGTGAAPRHYSGGIFNGECETDFA
SCGYNCCEBGTGYWVKNKSGENGEMRKRDVDAPQGWYYN"
complement(29643..30682)
```

```

/genes="At2g27410"
/notes="F10A12.21; predicted by genefinder"
complement(join(<29643..30233,30344..>30682))

mRNA
Query Match          9.8%; Score 117.2; DB 8; Length 90341;
Best Local Similarity 47.8%; Pred. No. 1.2e-16;
Matches 435; Conservative 0; Mismatches 468; Indels 7; Gaps 3;

QY 108 AGAAGATTCCAAATGTACAGAGATATAGAGAGATTGTTCTTCACTTCCCTCGAATGA 167
Db      |||
QY 80357 AAACGATGACTTAAGCGAAGAAACAAGACTCTCATCTCTTCACTACCTTCAGACAAAGA 80416
Db      |||
QY 168 TTATTGGGGGATACCATGAGTTGTACAGAGGATTGTCGAATGGGATATCTTGATCC 227
Db      |||
QY 80417 TTCCACTGGGATAAATGTTTGAAGTACCAAGGATGTTGTATACTCACCACITTCCTCA 80476
Db      |||
QY 228 TGGTATCATGGCTTTTGAAGATTAATTTCAAGGCTCGAGACGGAATATCTCTAGGAC 287
Db      |||
QY 80477 AGCTGTCTCAATTTCCAGAAAAAATTTCAAGCCTCAAGACCAATATCATCGTTGCTTC 80536
Db      |||
QY 288 TCTTCAAAGGCTGGACGACATGGACGAAGGACCTGACGTTTGGCATCTCTACATGATG 347
Db      |||
QY 80537 GTTCCCTTAATGCGGACGACTGGCTCAAGGCATTTACATTTCTACTAGTCATGATC 80596
Db      |||
QY 348 TGTAAACACCCATCATCACCGACACATCCACTTTTGTCTTCAACCCCTCATTCGTGTG 407
Db      |||
QY 80597 AAACACCCCTTCATGATCATCATCTCCCTTCTTTCAACAATCCGACGCTTCTGTG 80656
Db      |||
QY 408 TCAAAATTTGGAGTATTTGTACATGGGTAGAGAAATACGATCGCAGACCTCGATATGTT 467
Db      |||
QY 80657 CCCCTAACT--TAAATGAATCTCAATTAATCTACTAGCGAAAAACCGGACCTGACCA 80713
Db      |||
QY 468 GAATGAATCCGAGGTGTTTTCGCGACACATCCCATCTCTTGTGTCGGCGCTGTG 527
Db      |||
QY 80714 CTATCGTCTCCGAGGCTGTTCTCGACACATCGCGTCTCATACGCTGCAAGAAGTCT 80773
Db      |||
QY 528 TTTGAAATCGGGAACAAAAATCATCAATATAAGCGCAACCGTAAGAGTAGTACATTTGTGTC 587
Db      |||
QY 80774 TAAGATTCTACTTTGAAGTGTGTATATTGCGAGCAATGAAGACACATTTGGTTTA 80833
Db      |||
QY 588 TTTTGGAAATTTGGCAATCTGATTAAACCGGACAAAGTTATTGGACCTCGAAAAGCGT 647
Db      |||
QY 80834 ATATTGGCATTTCTTTCTAATGATGATAAGCTAAAAAACAGTTTTCGAGGAAACGTT 80893
Db      |||
QY 648 TGATATCTTCGATCGGGAATCTCTTTTGTGGACCGGAATGGAATTTCCAAGCGGAGTT 707
Db      |||
QY 80894 GTCTTTTGTNAGGAGTCAACTTTTTTGGCCCTTTTGGACCCAGTCTTTAAGCTACTG 80953
Db      |||
QY 708 CACCAATCGCGGCTCTACTAATTCAAACTTGCTATTGTTGAGTTACGAAGAAATGTTTGA 767
Db      |||
QY 80954 GAGAAGAGCTTGGAAAGATCCAAATCATGTCTTTTCATGAGGTTTGAGGAGATGAAGC 81013
Db      |||
QY 768 GAAGCCAGTTGAAAATGTGAAGAGCTAGCTGAGTTTCATGGGATGCTGGGTTCAAGACGA 827
Db      |||
QY 81014 AGAACCTCATGAGCAGATCAAGAGACTTGCTGAGTTC-TTAGATTCTCTCTTACTAAGGA 81072
Db      |||
QY 828 TGAGGAGAAACAAGGATGTTGTGATGAGATAGTTAAACTTTGTAGCTTCGACAACTCGAA 887
Db      |||
QY 81073 AGAAGAGAGAACGGATTGGTGGATGAGATATAGATCTTTGCTCTCTACGTAATCTAAG 81132
Db      |||
QY 888 GAATCAACAGGTGAACAAAAACCGATCAAGCTACAATTCGAAATTCGACAAACAGCATTT 947
Db      |||
QY 81133 CAGTTTGGAGATCAATAAGACCGGAAATTTATGTGAAGGTAGA--GATAACAAAACATT 81189
Db      |||
QY 948 CTTAGGAAAGGTGAGGTTGAGAGATTGGGCAAACTATCTACCTCGGAAATGATTAGAA 1007
Db      |||
QY 81190 TTTCCGTAAGGAGAGAGTCGCTGATTGGAAGAACTATTTTGACTCTCTGAAATGGAGAACAA 81249
Db      |||
QY 1008 ACTGGAGACG 1017
Db      |||
QY 81250 AATGACATG 81259
Db      |||
```

```

RESULT 13
LOCUS AC005824/c 118196 bp DNA linear PLN 11-MAR-2002
DEFINITION Arabidopsis thaliana chromosome 2 clone F15K20 map B68, complete
sequence.
ACCESSION AC005824
VERSION AC005824.3 GI:20197449
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsie.
REFERENCE 1 (bases 1 to 118196)
AUTHORS Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M.,
Shen,M., Ronning,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
JOURNAL Unpublished
TITLE 2 (bases 1 to 118196)
AUTHORS Lin,X.
JOURNAL Direct Submission
SUBMITTED (09-MAR-2000) The Institute for Genomic Research, 9712
MEDICAL CENTER DR., ROCKVILLE, MD 20850, USA
REFERENCE 3 (bases 1 to 118196)
AUTHORS Town,C.D. and Kaul,S.
JOURNAL Direct Submission
SUBMITTED (27-FEB-2002) The Institute for Genomic Research, 9712
MEDICAL CENTER DR. ROCKVILLE, MD 20850, USA, cdton@tigr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6598490.
FEATURES
source
1..118196
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="B68"
/clone="F15K20"
misc_feature
1..13541
/notes="overlap with BAC clone T1E2
(AC006929:81344..94884)."
complement (6251..6271)
/rpt_family="AT_rich"
repeat_region
10247..10271
/rpt_family="AT_rich"
gene
11482..12426
/genes="At2g27870"
/notes="F15K20.3"
<11482..>12426
/mRNA
/genes="At2g27870"
CDS
11482..12426
/genes="At2g27870"
/codon_start=1
/product="putative reverse transcriptase"
/protein_id="AAM15081.1"
/db_xref="GI:20197456"
/translation="MQQVLTNAERRRRHLSDSICQICKGAETIIHILRDCPAMEG
IWRLPVAKRREFFQSLEWLFANLGRRTKCTSTWTLFALSILWAWKRCGNIF
GVQDKDRVRFKDLARETSMHVIVRTLSCGHERVELIAWSKPEEGWKLNTDG
ASGNPLGASAGVLDDEGAWRGGFALNIGVCSAPLAELWGVYGLYIAWERRVRL
EIVDSSEIVVGFELKIGINEVHPLSLFLVRLCHDFISRDWRVIRSHVYREANRLADGLAN
YAFSLPLGFHSLSLVPDSLRFILDDTSGATVSRQVRM"
12779..15261
/genes="At2g27860"
/notes="F15K20.4; supported by full length cDNA:
Ceres:28205"
mRNA
join(12779..13109,13268..13360,13462..13569,13849..13963,
14039..14172,14261..14362,14472..14564,14654..14789,
14881..15261)
/genes="At2g27860"
CDS
join(12834..13109,13268..13360,13462..13569,13849..13963,
14039..14172,14261..14362,14472..14564,14654..14789,
14881..14993)
/genes="At2g27860"
/codon_start=1
```


Qy 288 TCTTCAAGGCTGGAACGACATGGAGGAGGCACTGACGTTGGCCATCTCTAACACGAGA 347
 Db 104136 GTTCCCTAAATGCGGACGACTGGCTCAAGGCACATTACATTTCTCACTAGTCATAGATC 104077
 Qy 348 TGTAAACCCATCATCAGGACACATCCACATTTTGTCTTCAACGCTCATTCGTGTGT 407
 Db 104076 AAAACACCCCTTCTCATGATCATCATCCCTTCTTCTTAAACAATCCGACGCTTCTGT 104017
 Qy 408 TCAAAATTTGGAGTATTGTGTACATGGGTAGAGAAATACCATCCAGACCTCGCATATGTT 467
 Db 104016 CCCTTAACT---TAAATGAATCTCAATTAATAGGAAAAACCGACCTGACCAAGTT 103960
 Qy 468 GAATGAATCCCGAGGTTGTTTGGCGGACACATCCCATCTCTTTGTGCGGCGTCTGT 527
 Db 103959 CTCATCGTCTCCGAGGCTGTTCTCGACACACATCGCGTCTCATACGCTGCAAGAAGTCT 103900
 Qy 528 TTGAAATCGGGAACAATAATCATCAATATAAGCCGCAACCGTAAGATGATCTTGTGTC 587
 Db 103899 TAAAGATTCTACTTTGTAAGTTGTGTATATTGAGGAACATGAAAGACACATTTGGTTTA 103840
 Qy 588 TTTTGGAAATTTGGCAATCTGATTAAACCCGACAGTATTGGACCTCGAAAAAGAGCGT 647
 Db 103839 ATATTGGCAATTTCTTTCTAATGATTAACGTAATAAACCAAGTTTCAGGAAACGTTTCA 103780
 Qy 648 TGATATCTTCGCATCGGGAATCTCTTTTGTGGACCGAATGGAATTTCCAAAGCGGAGTT 707
 Db 103779 GTCTTTTGTAAAGGAGTCAACTTTTGTGGCCTTTTGGGACCAAGTCTTAAAGTACTG 103720
 Qy 708 CACCAATCGCGGCTCTACTAATTCAAACTTGCTATTGTTGAGTTACGAAAGATGTTAGA 767
 Db 103719 GAGAAGAAGCTTGAAGATCCAAATCATGTCCTTTTCATGAGGTTTGAGGAGATGAAGC 103660
 Qy 768 GAAGCCAGTTGAAATGTGAAGAGCTAGCTGAGTTCATGGATGTTGGTTTACAGACGA 827
 Db 103659 AGAACCTCATGACAGATCAAGAGACTTCTGAGTTC-TTAGATTCTCTTACTAAGGA 103601
 Qy 828 TGAGGAGAAACAAGGATTTGTGTAGATAGTGTAAACTTTGTAGTCTTCGACAATCTGAA 887
 Db 103600 AGAAGAAGAACGATTTGGTGGATGAGATTATAGATCTTTGCTCTCTAGCTAATTAAG 103541
 Qy 888 GAATCAACAGGTGAACAAAACGATCAAGTACAAATTCGAAATTCACACACAGCATTT 947
 Db 103540 CAGTTTGGAGATCAATAAGACCGGAAAATTTATGTAAGGTAGA---GATAACAAAACAT 103484
 Qy 948 CTTTCAGGAAGCTGAGGTGAGAGATTGGGCAAACTATCTAACGTCGGAATATGATTAAGAA 1007
 Db 103483 TTTCCTGAAGAGAGATCGGTGATTGGAAGAACTATTGACTCTGGAATGGAGAACAA 103424
 Qy 1008 ACTGAGAGC 1017
 Db 103423 AATCGACATG 103414

RESULT 14
 AF375458
 LOCUS AF375458 1179 bp mRNA linear PLN 07-FEB-2002
 DEFINITION Arabidopsis thaliana At2g03760/F19B11.21 mRNA, complete cds.
 ACCESSION AF375458
 VERSION AF375458.1 GI:14030734
 KEYWORDS FLI CDNA.

ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1179)
 Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Shinn, P.,
 Ban, J., Bowser, L., Carninci, P., Chung, M.K., Goldsmith, A.D.,
 Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G.,
 Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,
 Narusaka, M., Nguyen, M., Palm, C.J., Pham, P.K., Quach, H.L.,
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,
 Toriumi, M., Yamada, K., Yu, G., Shinozaki, K., Davis, R.W.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Theologis, A. and Ecker, J.R.
 Arabidopsis cDNA clones
 2 (bases 1 to 1179)
 Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Shinn, P.,
 Ban, J., Bowser, L., Carninci, P., Chung, M.K., Goldsmith, A.D.,
 Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G.,
 Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,
 Narusaka, M., Nguyen, M., Palm, C.J., Pham, P.K., Quach, H.L.,
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,
 Toriumi, M., Yamada, K., Yu, G., Shinozaki, K., Davis, R.W.,
 Theologis, A. and Ecker, J.R.

TITLE
 JOURNAL

Submitted (01-MAY-2001) Salk Institute Genomic Analysis Laboratory
 (SIGNAL), Plant Biology Laboratory, The Salk Institute for
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
 USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA : RIKEN
 Arabidopsis Full-length cDNA) : Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.

COMMENT

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Shinn, P., Chen, H.,
 Cheuk, R., Kim, C.J., Koesema, E., Meyers, M.C., Tracy, S.E., Ban, J.,
 Bowser, L., Chung, M.K., Goldsmith, A.D., Jones, T., Karlin-Neumann, G.,
 Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M.,
 Palm, C.J., Pham, P.K., Quach, H.L., Sakano, H., Southwick, A.,
 Tang, C.C., Toriumi, M., Yamada, K., Yu, G., Davis, R.W., Theologis, A.,
 and Ecker, J.R.

Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
 contributed equally to this work as PIs.

FEATURES

source

1..1179
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosomes="2"
 /clone="RAFL05-01-D08(R12477)"
 /notes="ecotype: Columbia"
 63..1043
 /notes="putative steroid sulfotransferase"
 /codon_start=1
 /product="At2g03760/F19B11.21"
 /protein_id="AAK53042.1"
 /db_xref="GI:14030735"

CDS

/translation="MSSSSVPAYLGDDELDTQETRALISSLPKEKGLVSEIYVEOGL
 WHTQALQGLICQKRFKAKSDIILVTPKSGTTWTKALVFNLRHFPVSSGNH
 PLLVTNPVLPVLEGVYSPDFSSLPRLMNTHTSHLSLSPVSSSKIVY
 CRNPKDMFVSLWHFGKLAPEETADYPIEKAVEAFCEGFIGGPFWDHILEYVYASRE
 NPNKLVFTYEELKQTEVMKRIAEFLGCFIEBEVEVRIKVLCSFSLNSLNVNKE
 GKLPNGIETKTPRKEIGGWRDITLSELAEBIDRTIEBKFKGSLKLFSS"

BASE COUNT 335 a 204 c 257 g 383 t
 ORIGIN

Query Match 9.6%; Score 114; DB 8; Length 1179;

Best Local Similarity 50.5%; Pred. No. 7.5e-16;

Matches 487; Conservative 0; Mismatches 435; Indels 42; Gaps 7;

Qy 141 GATTGTTTCTTCACTCCCTCGAATGATTATTGCGGGGATACCATGAGTTGTACAAGG 200

Db 131 GATCTCTTCTCTCTCTAAAGAGAAAGGTTGTTAGTGAAGTAATATATGATTTCAAGG 190

Qy 201 ATTTGGCAATGGATATCTTCTACTGCTATCATGGCTTTTCAAGATAATTTCAAGGC 260

Db 191 ACTTTGGCACACACAGCTATTTTACAAGGAATCTTGATCTGCCAAAACGCTTTGAAGC 250

Qy 261 TCGAGAGACGAGCATTTATCTTACGACTCTTCCAAAGGCTGGAAACGACATGGACGAGG 320

Db 251 TAAAGATTCCGACATTATCTCTCGTCACTAATCTCTAAATCAGGTACCATCTTGGTTAAAGC 310

Search completed: June 10, 2003, 18:20:39
Job time : 3140 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 17:10:11 ; Search time 315 Seconds
(without alignments)
8521.848 Million cell updates/sec

Title: US-09-854-122-15
Perfect score: 1192
Sequence: 1 acgggggaataactggaat.....gttaaaaaaaaaaaaaaa 1192

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

```
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1192	100.0	1192	24	z. marina cDNA clo
2	128.6	10.8	1212	21	Arabidopsis thalia
3	127.2	10.7	1154	21	Arabidopsis thalia
4	112.4	9.4	1205	21	Arabidopsis thalia
5	110.6	9.3	1043	21	Arabidopsis thalia
6	103.2	8.7	1281	21	Arabidopsis thalia
7	102.4	8.6	1273	21	Arabidopsis thalia
8	100.8	8.5	1077	22	Hydroxyjasmonic ac
9	100.8	8.5	1270	21	Arabidopsis thalia

10	94.4	7.9	1347	21	AAC37555	Arabidopsis thalia
11	89.2	7.5	1041	22	AAF29178	Hydroxyjasmonic ac
12	79.8	6.7	1160	21	AAC32748	Arabidopsis thalia
13	55.4	4.6	344	24	ABQ85922	Arabidopsis thalia
14	48.4	4.1	7303	24	ABK31302	DNA transcription
15	48.2	4.0	10467	24	ABK28453	Signal transductio
16	48	4.0	11787	24	ABL92243	Chemically treated
17	47.2	4.0	12578	22	AAS46659	Tumour suppressor
18	46.8	3.9	24259	22	AAS46692	Tumour suppressor
19	46.2	3.9	3533	23	ABL08490	Drosophila melanog
20	45.8	3.8	6112	24	ABK33981	Human DNA for stag
21	45.6	3.8	6123	24	ABL33036	Human immune syste
22	45.4	3.8	743	24	ABQ65471	Arabidopsis thalia
23	45.4	3.8	40324	24	ABQ67150	Human angioogenesis
24	45.2	3.8	7455	24	ABL33759	Human immune syste
25	44.8	3.8	6247	24	ABL32274	Human immune syste
26	44.8	3.8	14861	24	ABL70554	Chemically treated
27	44.8	3.8	14861	24	ABL33195	Human immune syste
28	44.8	3.8	14861	24	AAS61203	Human gene regulat
29	44.8	3.8	17580	24	ABL54312	Human gene regulat
30	44.8	3.8	17580	24	ABN80013	Human chemically m
31	44.8	3.8	50000	24	ABL56202	AmEPV genome fragm
32	44.4	3.7	6107	24	ABL70390	Chemically treated
33	44.4	3.7	6107	24	AAS61342	Human gene regulat
34	44.4	3.7	6107	24	ABK31431	Signal transductio
35	44.2	3.7	17389	24	ABL33414	Human immune syste
36	44	3.7	7511	24	ABL33283	Human immune syste
37	44	3.7	19787	24	ABL33451	Human immune syste
38	44	3.7	23695	24	ABQ66981	Human angioogenesis
39	43.8	3.7	6095	22	AAS46310	Tumour suppressor
40	43.8	3.7	6095	24	ABL70150	Chemically treated
41	43.8	3.7	6095	24	ABL32361	Human immune syste
42	43.8	3.7	6095	24	ABL34475	Human metastasis a
43	43.6	3.7	9021	22	AAS46325	Tumour suppressor
44	43.6	3.7	20579	24	ABQ67074	Human angioogenesis
45	43.4	3.6	8099	24	ABL33723	Human immune syste

ALIGNMENTS

RESULT 1
AAS20863
ID AAS20863 standard; cDNA; 1192 BP.

XX AC AAS20863;

XX DT 09-APR-2002 (first entry)

XX DE Z. marina cDNA clone encoding protein containing sulfotransferase.

XX KW Plant; transgenic; marine eelgrass; zosteric acid biosynthesis;

XX KW saline-resistance; anoxia-resistance; anti-fouling genetic trait;

XX KW marine vascular plant; sulphated phenolic compound; Zostera marina;

XX KW sulfotransferase; ST; enzyme; gene; ss.

XX OS Zostera marina.

XX FH Key Location/Qualifiers
XX CDS 1..1192
XX FT /*tag= a
XX FT /partial
XX FT /product= "Protein containing sulfotransferase"

XX FT /note= "This sequence lacks both start and stop codons"
XX FT /transl_except= (pos:34..36, aa:Xaa)
XX FT /transl_except= (pos:1042..1044, aa:Xaa)
XX FT /transl_except= (pos:1060..1062, aa:Xaa)
XX FT /transl_except= (pos:1078..1080, aa:Xaa)
XX FT /transl_except= (pos:1087..1089, aa:Xaa)
XX FT /transl_except= (pos:1120..1122, aa:Xaa)
XX FT /note= "Xaa= Stop codon"

XX PN WO200185971-A2.

PR 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 23-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134356.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134219.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141484.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

```
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0159117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161320.
PR 28-OCT-1999; 99US-0161392.
PR 28-OCT-1999; 99US-0161593.
PR 29-OCT-1999; 99US-0162142.

Query Match      10.8%; Score 128.6; DB 21; Length 1212;
Best Local Similarity 48.4%; Pred. No. 2.8e-21;
Matches 452; Conservative 0; Mismatches 469; Indels 12; Gaps 3;

QY 88 AAGAATGACGACGAGAGAGAGATTCACAAATGTACAGAGATATAGAGAGATGTGT 147
DB 61 AAGGAGAAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 148 TCTTCACTTCCCTCGAATGATTATTTGGGGGATACCATGAGTTGTACAGGATTTGG 207
DB 121 TCACCTTCTTCAGACATAGATTGCTCAGGGACCAAGTTGTACAGTACCAAGGATGTGG 180
QY 208 CAAATGGGATATCTTGTACCTGTGTATCATGGCTTTCGAAGATATTTCAAGGCTCGAGAG 267
DB 181 TAGCACAAGATATTTCTCAAGCAATCTCAATTTCAACAAAATCTTCAGCCACAGAA 240
QY 268 ACGGACATATTCTTACGACTCTTCCAAAGCTTGGACGACATGGACGAGGCACTGACG 327
DB 241 ACGGATATAATTGTTGCTTCTTCCCAATCGGTCAGGACTTGGCTCAAGGCACTCACA 300
QY 328 TTTGCCATCTCAACAGAGATTTAACCCCATCATCCGACACATCCACTTTTGTTC 387
DB 301 TTGCACTTCGCGCAAGA-----TCAAAACACACTTCAGACCAATCATCTCTGCTAACT 354
QY 388 TTCACACCTCATCTCGTGTGTTCAAATTTTGGAGTATTTGTACATGGGTAGAGAAATACG 447
DB 355 CATATCTCATGAGCTAGTGCCTGACTCTGAGCTCGATCTTTATCTCAAAAGCTCGAAA 414
QY 448 ATGCCAGACTCGATATGTTGAATGAATCGCCGAGGTTGTTGCCGACACATCCCATAC 507
DB 415 CCGGATTGACCAAGTTGCCATCATCATCTCGAGATTTTCTCAACCCACATGTCCTTT 474
QY 508 TCTTTGTTGCGGCGTCTGTTTGAATTCGGACACAAAATCATCATATATAAGCCGCAAC 567
DB 475 GATGCGCTTAAAGTACCGTTGAAAGAGTCTCTCTTTCGCAAGATCGGTGATCGGAGAAC 534
```

```
QY 568 CGTAAGAGTACATTTGTGTCTTTTGGAAATTTGGCAATCTGATTAACCCCGACAAGTTA 627
DB 535 GTGAATGACGTATTGATATCACTTTTGGTGTTCGAAAACTCCATGAGTGGAGAAAACAT 594
QY 628 TTGACCTCGAAAAGAGCGTTGATATCTTCGCATCGGGAATCTCTTTTGTGGACCGAA 687
DB 595 TTAAGTCTCGAGGCTTTTGTTCGAGTCTTTATGTAGCGGAGTTAACTTTATCGGTCCTTTA 654
QY 688 TGG--AAATTTCCAAAGCGAGTTTCAACAAATCGCGGCTCTACTAATTTCAAACCTTGCTATTG 744
DB 655 TGGGAAAATGTTTGGCTATTGGAGGAGAGCTTGGAGATCCTTAAGCATGTGCTTTTC 714
QY 745 TTGAGTTACGAAGAAATGTTAGAGAAGCCAGTTGAAAATGTGAAGAAGCTAGCTAGTTC 804
DB 715 TTGAGGTACGAGGAGTTGAAGACGAGCTCTCGTGTGCAAAATCAAGAGACTTTGCAGAGTTC 774
QY 805 ATGGGATGTGGTTTCACAGACGATGAGGAGAAACAAGGATTTGTGTAGATAGTATAA 864
DB 775 TTAGATTGTCTCATTAAGAAACCTTAGCGGTTGGAGATCAACAAAACAGG---AAGCTTGTTCG 834
QY 865 CTTTGTAGCTTCGACAATCTGAAGAATCAACAGGTGAACAAAAACGGATCAAGCTTACAAT 924
DB 835 CTTTGTCTCTAAGAAACCTTAGCGGTTGGAGATCAACAAAACAGG---AAGCTTGTTCG 891
QY 925 TCGAAAATCGAACAAAGCATTTTTCAGGAAAGGTGAGGTGAGAGATTTGGGCAACTAT 984
DB 892 GAAGGAGTAAGTTTCAAGAGTTTTTTCGTAAGGGGAAAGTTGGTGTATTGGAAGAGTTAT 951
QY 985 CTAACGTGGAAATGATTAAAGAACTGGAGAG 1017
DB 952 ATGACTCTCTGAATGGAACAAAATCGACATG 984

RESULT 3
AAC37672
ID AAC37672 standard; DNA; 1154 BP.
XX
AC AAC37672;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 18246.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
```

PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140682.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144326.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145090.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.

PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161320.	
PR	28-OCT-1999;	99US-0161392.	
PR	28-OCT-1999;	99US-0161393.	
PR	28-OCT-1999;	99US-0162142.	
Query Match 10.7%; Score 127.2; DB 21; Length 1154;			
Best Local Similarity 50.8%; Pred. No. 6e-21;			
Matches 468; Conservative 0; Mismatches 433; Indels 21; Gaps 6;			
Qy	101	AGAAGGAAGAGATTCCAAATGTACAAGAGATATAGAGAGATTGTTCTTCACTTCCTT	160
Db	130	ACATGAAGACGACAACTTAGTCAAGAAACCAAGACTTGTATCACTTCTCTACCTTCAG	189
Qy	161	CGRATGATTATGGGGGATACCATGAGTTGTACAAGGGATTGTCGCAATGGGATATC	220
Db	190	ACAAAGATTTCATGGGTTATGGTCTCTACAACCTACAAGGTTGTTGGTACTATCCAAACA	249
Qy	221	TTGTACCTGGTATCATGGCTTTTCAAGATATATTTCAAGGCTCGAGAGACGGACATTATCC	280
Db	250	CACCTCAAGCCGTTCTTGACGTCCAAAACACACTTCAAGCCACGAGATACCTGATATAATCC	309
Qy	281	TTACGACTTCTTCCAAAGCTGGAAAGCATGAGAGAGGAGGACTGACGTTTGGCCATCTAA	340
Db	310	TCGCTTCTTTCGCCAAAGTGGAAACCACTTGGCTCAAAATCCCTAAATTTTCGCTGTGTAC	369
Qy	341	CACGAGATGTTAAACCCATCATCACGACACATCCACTTTTGTCTTCAACCCCTCATTT	400
Db	370	ATAGAGAAAGTACCGGGAACCCCTCAACACATCTTGTCTTTACAAACCCCTCATG	429
Qy	401	CGTGTGTTCAAAATTTGGAGTATTTGTACATGGGTAGAGAAATACGATGCGGACCTCG	460
Db	430	ACCTTGTCCTTCTTGAGGTTGAGTTATACGCTA---ATAGCCAAATTCGGGATCTCG	486
Qy	461	ATATGTTGAATGAATCGCCGAGTTGTTGGCGACACATCCCATCTTTGTTGGCGG	520
Db	487	CAAAAGT---ATTCTTCTCTATGATCTTTTCTACACACATGCCTTACAAGCATTTGGTG	543
Qy	521	CGTCTGTTTTCGGAACGGAACAAATATCAATATAAGCCGCAACCGTAAGAGTACAT	580
Db	544	AAGCCACCAAAAGCTTG---CAAAACCGTATATGTTGTAGAGGTATCAAGATACGT	600
Qy	581	TTGTGCTCTTTTGGAAATTTGGCAATCTGATTAACCCCGCAAGTT-----ATTGACC	634
Db	601	TTGCTCCGGCTGGCATTTATGAACATGTTGTCATCGCAACCAAGATGATCAAGCCACTT	660
Qy	635	TGAAAAGAGCGTTGATATCTTCGATCGGNACTCTCTTTTGTGGACCGGATGGAAAT	694
Db	661	TTGAGCTCATGTTTGATGCTTATTTAGAGAGGTTCTCTTATATGGACCTTATTGGGAAC	720
Qy	695	TCCAAGCGGAGTTTCAACCAATGGGGCTCT---ACTAATTCAAACTTGTCTATTGTTGAGTT	751
Db	721	ATGTATTGAGCTATTGGAAAGGAGCTTGGAAAGCAAGGAGATGTTCTTTTCATGAAGT	780
Qy	752	ACGAAGAAATGTTAGAGAGCCAGTTGAAATGTGAAGAGCTAGCTAGTTCATGGAT	811
Db	781	ACGAAGAGATAATTGAGGAGCCCTCGTGTTCAGTCAAGAGACTCGCCGAGTTCTTTGGAAT	840
Qy	812	GTGGGTTACACAGCATGAGGAGAACAGGATTTGTTGATGAGATAGTTAAACTTTGTA	871
Db	841	GTCCATTTCACGAAGGAAGAAAGTGGATTCGGTGGAGGATCTTGAAGTTGTGTA	900

Qy	872	GCTTCGACAATCTGAAGAAATCAACAGGTGAACAAAAACCGATCAAGCTACAAATTCGAAAA	931
Db	901	GTTTACGAAATTTAAGCAATTTGGAGGTTAATAAGAATGGGACCAACGAGAAAT---GGTG	957
Qy	932	TCGACAACAAGCATTTCTTCAGGAAAGGTGAGGTTGAGAGATTGGGCAAACTATCTAAAGT	991
Db	958	TAGATTCTCAGGTGTTCTTTAGGAAAGGTGAAGTTGGTGATTGGNAGAAATCATCTTACGC	1017
Qy	992	CGGAAATGATTAAAGAACTGGA	1013
Db	1018	CACAAATGGCGAAACCTTTGA	1039
RESULT 4			
AAC37521	AAC37521 standard; DNA; 1205 BP.		
ID	AAC37521	(first entry)	
XX	AAC37521;		
AC	17-OCT-2000	Arabidopsis thaliana DNA fragment SEQ ID NO: 17688.	
DT	XX	Hybridization assay; genetic mapping; gene expression control;	
XX	XX	protein identification; signal transduction pathway;	
DE	XX	metabolic pathway; promoter; termination sequence; ss.	
XX	XX	Arabidopsis thaliana.	
KW	XX	Arabidopsis thaliana.	
XX	XX	EP1033405-A2.	
PN	XX	06-SEP-2000.	
PD	XX	25-FEB-2000; 2000EP-0301439.	
XX	XX	25-FEB-1999; 99US-0121825.	
PR	XX	05-MAR-1999; 99US-0123180.	
PR	XX	09-MAR-1999; 99US-0123548.	
PR	XX	23-MAR-1999; 99US-0125788.	
PR	XX	25-MAR-1999; 99US-0126264.	
PR	XX	25-MAR-1999; 99US-0126785.	
PR	XX	01-APR-1999; 99US-0127462.	
PR	XX	08-APR-1999; 99US-0128234.	
PR	XX	16-APR-1999; 99US-0128714.	
PR	XX	16-APR-1999; 99US-0129845.	
PR	XX	19-APR-1999; 99US-0130077.	
PR	XX	21-APR-1999; 99US-0130449.	
PR	XX	23-APR-1999; 99US-0130510.	
PR	XX	23-APR-1999; 99US-0130891.	
PR	XX	28-APR-1999; 99US-0131449.	
PR	XX	30-APR-1999; 99US-0132048.	
PR	XX	30-APR-1999; 99US-0132407.	
PR	XX	04-MAY-1999; 99US-0132484.	
PR	XX	05-MAY-1999; 99US-0132485.	
PR	XX	06-MAY-1999; 99US-0132486.	
PR	XX	06-MAY-1999; 99US-0132487.	
PR	XX	07-MAY-1999; 99US-0132863.	
PR	XX	11-MAY-1999; 99US-0134256.	
PR	XX	14-MAY-1999; 99US-0134218.	
PR	XX	14-MAY-1999; 99US-0134219.	
PR	XX	14-MAY-1999; 99US-0134221.	
PR	XX	14-MAY-1999; 99US-0134370.	
PR	XX	18-MAY-1999; 99US-0134768.	
PR	XX	19-MAY-1999; 99US-0134941.	
PR	XX	20-MAY-1999; 99US-0135124.	
PR	XX	21-MAY-1999; 99US-0135353.	
PR	XX	24-MAY-1999; 99US-0135629.	
PR	XX	25-MAY-1999; 99US-0136021.	
PR	XX	27-MAY-1999; 99US-0136392.	
PR	XX	28-MAY-1999; 99US-0136782.	
PR	XX	01-JUN-1999; 99US-0137222.	
PR	XX	03-JUN-1999; 99US-0137528.	
PR	XX	04-JUN-1999; 99US-0137502.	

PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139455.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142330.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 13-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145313.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 13-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 16-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 9.4%; Score 112.4; DB 21; Length 1205;
 Best Local Similarity 47.5%; Pred. NO. 2.3e-17;
 Matches 437; Conservative 0; Mismatches 471; Indels 12; Gaps 3;

QY 101 AGAAGGAAGAGATTCCTCAAAATGTACAGAGATATAGAGAGATTGTTCTTCACTTCCT 160
 DB 62 AGGAAGAAGAAACCAAAAGTGAAGATTCCTCAAAAGTGTGATCTCTTCACTTCCTTCAG 121
 QY 161 CGAATCATTTATGGGGGGTACCATCAGGTTGTACAAAGGATTTTGGCAAAATGGGATATC 220
 DB 122 ACATAGATTGCTCTGGGACCAAGTTGTACAGATATCAAGGATGTTGGTACGATAAAGACA 181
 QY 221 TTGTACCTGGTATCATGGCTTTTCGAAGATAATTTCAAGGCTGAGAGACGACATTATCC 280
 DB 182 TTCTCCAAGCAATCTCTCAAAATCTTCAAGGATCTTTCAGGACCAAGAAACCGATATAATTG 241
 QY 281 TTACGACTCTTCAAAAGGCTGGAACGACATGAGGAGGACCTGAGCTTGGCATCTTAA 340
 DB 242 TTGCTTTCTTCCCAAAATCAGGTACGATTTGGTCAAGGACCTACATTCGCACTCGCTC 301
 QY 341 CAGGAGATGTTAAACCCCAATCATCACCGACACATCCACTTTTGTCTTCAACCTCATTT 400
 DB 302 AAAGA-----TCAAACATACCTTCAGAAATCATCTCTGCTAACTCATAATCCTCATG 355
 QY 401 CGTGTGTTCAAAATTTGGAGTATTTGTACATGGGTAGAGAAATATAGTCCAGACCTCG 460
 DB 356 AGCTAGTGGCGTACCTCGAGCTCGATCTTTATCTCAAAAGCTCGAAACCGGATATGTCCA 415
 QY 461 ATATGTTGAATGATCGCCGAGCTGTTTCCGAGACACATCCCATCTCTTGTGTCGGG 520
 DB 416 AGTTACCATCATCTCTCCGAGATTTGTTCTCAACCCACATGCTTTCGATCGCCTTAAG 475
 QY 521 CGTCTGTTTGAATCGGGAACAAATATCAATATAAGCCGCAACCGTAAAGATACAT 580
 DB 476 TACCAATGAGGAGACTCTCTTCAAGATAGTGTGTGTCAGGAAACGTAAGACGCTGT 535
 QY 581 TTGTGCTTTTGGAAATTTGGCAATCTGATTTAACCCCGACAGTTATTTGGAACCTCGAA 640
 DB 536 TGGTATCACCTTGGTGTTCGAAACCTCCATTAGTGAGAAACAAATTTAAGTCTCGAGG 595
 QY 641 AGACGTTGATATCTTCGATCGGATCTCTTTTGTGACCGGAATGG---AATTTC 697
 DB 596 CTTTGTTCGAGTCTTTATGATAGCGGAGTTAACTTATCGGTCCTTGTGGGAAATGTGT 655
 QY 698 AAGCGGAGTTTCAACCAATCGCGCTCTACTAATTCAAACTTGTCTTGTGAGTTACGAAG 757
 DB 656 TAGCTATTGGAGAGGAAGCTTGAAGATCTTAAGCATGTCTTCTTCTGAGGTACGAGG 715
 QY 758 AATGTTAGAGAACCCAGTTGAAATGTGAAGAGCTAGCTGAGTTCATGGATGTCGGT 817
 DB 716 AGTTGAAGACGGAGCCTCGTGTGCAAAATCAAGAGACTTGCAGAGTCTTAGATTTTCCAT 775
 QY 818 TCACAGAGCTAGGAGAGAAACAAGGATTTGTTGATGAGATAGTTAAACTTTGTAGCTTCG 877
 DB 776 TCACAAAGGAAGAAGATAGTGGAGGTGTAGCAAGATCTTGGAACTTTGTTCTCTAA 835
 QY 878 ACAATCTGAAGATCAACAGGTGAACAAAAACGATCAAGTCAAAATCGAAAAATCGACA 937
 DB 836 GAAACCTTAGCGGTTTGGAGATCAACAAAAACAGG---AAGCTTGTGCGAAGGAGTAAGTT 892
 QY 938 ACAAGCATTTCTCAGGAAAGGTGAGGTGAGAGATTTGGCAAACTATCTAACGTCGGA 997
 DB 893 TCAAGAGTTTTTTCGTTAAAGGGAAGTTGGTGTGGAAGAGTTATATGACTCTCTGAAA 952
 QY 998 TGATTAAAGAACTGGAGACG 1017
 DB 953 TGGAAACAAATCGACATG 972

RESULT 5

AAC42382

ID AAC42382 standard; DNA; 1043 bp.

XX

AC AAC42382;

XX

DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 35341.
 XX DE Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 XX PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.

[illegible]

Wed Jun 11 10:48:27 2003

us-09-854-122-15.rng

```
QY 438 AGAAATACAGATCCAGACTCGATATGTTGAATGAA-----TCGCCAGGTTGTTGCCGG 494
Db 423 ----TACGAGTCCCGAGATTTCGATTTTCCAGTTTGCCTTCTCCAGACTGATGAACAC 478
QY 495 ACACATCCCATCTCTTTGTCGCCGCTGTTTGAATCGGAAACAAAATCATCAA 554
Db 479 GCACATATCGCATCTTTGCGTCCCGAGTCTGTTAAGAGCTGCTGTTGAAGATTGTTGTA 538
QY 555 TATAAGCCGCAACCGTAAGAGTACATTTGTTGCTTTTGGAAATTTGGCAATCTGATTA 614
Db 539 TTGTTGTAGAACCCCTAAGACATGTTTGTGCTTATGCGATTTTGGGAAAAGCTAGC 598
QY 615 CCCGCAAGATTATGGACC-----TCGAAAAGAGGTTGATATCTTCGCATCGGGAAT 668
Db 599 TCCTGAGGAACCGCGGATATCTATCGAAAAGCGGTTGAAAGCGTTTGTGAAGGGAA 558
QY 669 CTCCTTTTGTGGACCGGAATGGAATTTCCAAAGCGGAGTTCCACCAATCGCGCGTCTACTAA 728
Db 659 GTTATAGGTGGACCTTTTGGGATCATATATTGGAGTACTGTTGCAAGCCGCGAGAA 718
QY 729 TTCAAACTTGCTATTTGTTGAGTTTACGAAGAAATGTTAGAAAGCCAGTTGAAAATGTGAA 788
Db 719 TCCGAACAAGTCTTGTGTTGTTACTTACGAGGAGCTAAAGAACGACACCGAAGTTGAGAT 778
QY 789 GAAGC---TAGCTAGTTTCATGGATGTTGAGTGTGTTACAGACGATCAGGAGAAACAAGGAT 845
Db 779 GAAGCGGATCGCGAGTTCTTGGAAATGTTGGAATGTTGTTAAGAAAGA-----826
QY 846 TGTTGATGAGATAGTTAAACTTTTGTAGTTTCGCAATCTGAAGAAATCAACAGGTGAACAA 905
Db 827 AGTGAAGAGATTGTGAAGTTGTGAGTTTGTAGTTTGAAGTTTAAAGTTTGAAGTTAACA 886
QY 906 AAACGGATCAAGCTACAAATTCGAAATTCGAAATTCGAAATTCGAAATTCGAAATTCGAA 965
Db 887 AGAAGGAAATTTGCAAAAT---GGAATAGAGACTTAAACTTCTTTTGAAGAAAGGAGAT 943
QY 966 GAGAGATTGGCAACTATCTAACGTCGGAATGATTAAAGAACTGGAGACGCGCGGAAA 1025
Db 944 TGGAGATTGGAGAGATCTTTGAGTGAGTCTATGTCAGAGGAATTTGATAGAACCATTTGA 1003
QY 1026 AATAATGAATCAGAGTAAAGCAATTTTATTCGT 1060
Db 1004 AGAGAAGTTTAAAGTTCTGCTCTTAAATTTCTT 1038

RESULT 6
AAC37716
ID: AAC37716 standard; DNA; 1281 BP.
XX AAC37716;
AC AAC37716;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 18414.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
```

PR	19-JUL-1999;	99US-01443325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144332.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	25-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 8.7%; Score 103.2; DB 21; Length 1281;		
Best Local Similarity 49.9%;		

```
QY 769 AAGCCAGTTGAAATCTGAAGAGCTAGCTGAGTTTCATGGATGTTGGTTTCACAGCGAT 828
Db 840 GATCCTTTACCGTACGTGAAGAGTCTGGCTGAGTTTATGGTCATGGATTACAGCCGAG 899
QY 829 GAGGAGAAACAAGGGATTGTTGATGAGATAGTATTAACCTTTGTAGCTTCGACAACTCTGAAG 888
Db 900 GAAGAGGAGAAAGTGTGTTGAGAAAGTGTGAATCTTTTCAGCTTCGAGAGCTTGAAG 959
QY 889 AATC-----AACAGTGAACAAAAACGGATCAAGCTACAAATTCGAAATCGCAAC 939
Db 960 AATCTTGAAGCTAACAAAGGGGAGAAAGACAGAGGATCGTCTGTGTTTACCGCAAT 1019
QY 940 AAGCAATTTTCAGGAAGCTGAGGTGAGAGATTGGGCAAACTATCTAACGTCGGAATG 999
Db 1020 AGCGCGTATTTTCAAGAAAGAAAGGTGGGAGATTGTTGCAACTATCTGACTCCGGAGATG 1079

RESULT 7
AAC47848
ID AAC47848 standard; DNA; 1273 BP.
XX
AC AAC47848;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55342.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PP 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135623.
PR 25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
19-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
22-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
```

```
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.
Query Match 8.6%; Score 102.4; DB 21; Length 1273;
Best Local Similarity 48.7%; Pred No. 5.9e-15;
Matches 441; Conservative 0; Mismatches 441; Indels 24; Gaps 5;

QY 96 GCAAGAGAAGGAAGATTCAAAATGTACAAAGATATAGAGAGATTTGTTCTTCACT 155
Db 167 GAAAGCGAAGAAGATGAAGGCTAAGCTGCGAGTTCCAAGAGATGTTGGATTCTCTCC 226
QY 156 TCCCTCGAATGATTATGGGGGATACCATGAGGTGTGTAAGGGATTTTGGCAATGG 215
Db 227 TAAGGAGAGAGATGGAACTCGTTACCTTTTCAAGGCTTTTGGTCCCAAGC 286
QY 216 ATATCTTGATCTGCTGATCATGGCTTCGAAGATTAATTTCAAGGCTCGAGAGACGACAT 275
Db 287 CAAAGAGATTCAAGCCATCATGTCTTTCCAAAACATTTCCAAATCCCTCGAAAACGACGT 346
QY 276 TATCCTTAGACTCTTCCAAAGGCTGGAACGATCGAGCAAGGCACTGACGTTTCCCAT 335
Db 347 CGTCTCGCCACCATACCTAATCCGTTACAACTGGCTTAAAGCTTTTAACTTTCAACAT 406
QY 336 CCTAACACGAGA-----TGTTAACCCCATCATCCGACACATCCACTTTTGTCTT 389
Db 407 CCTTAACCGTCAACCGTTGATCCGGTTGCTCGAGTACCAACCACTCTTTTCACTTC 466
QY 390 CAACCTCATCTGCTGTTCAAAATTTGGAGTATTGTTACATGGGTAGAGAAAATACGAT 449
Db 467 CAACCTCATGACCTTGTTACCTTTCTTGAGTACAGCTTTTACGCCAACGGAGATGTTCC 526
QY 450 GCCAGACCTCGATATGTTGAATGAAATCGCCGAGGTGTTTGGCGGACACATCCCATPCTC 509
Db 527 CGATCTCTCGGCTAGCCAGT-----CCAAGAAGTTTCGCAACCCACTTACCCTGTCGG 580
QY 510 TTTGTTGCGGCGCTGTTTGAATCGGGAACAAAATCATCAATATAAGCGGCAACCG 569
Db 581 TTCCTTAAAGGAACGATCGAGAAACCCGCTGTAAGGTGCTGTTCTTGTGCGGAACCC 640
QY 570 TAAGAGTACATTTGTTGTTTGGAAATTTGGCAATCTGATTAAACCCGA-----CAA 623
Db 641 GTTTCACACATTCATCTCTTCTGTCGCAATTACCAACAAACATCAATCCGAGTCAGTGAG 700
QY 624 GTTATTGGACCTCGAAAAGAGCGTTGATATCTTCGCATCGGGAATCTCTTTTGTGGACC 683
Db 701 CCCAGTCTTGTAGACCAAGCTTTTGTATCTGTTATTCGCGGGAGTGATCGGTTTGGCCC 760
QY 684 GGAATGGAATTTCCAAGCGGAGTTCAACCAATCGCGGCTTACTAATTCA---AACTTGCT 740
Db 761 GTTTTGGGAACACATGTTGGGATCTGGAGAGAGAGCTTTGAAGAGACCAGAAAGTCTT 820
QY 741 ATTGTTGAGTTACGAAGAAATGTTAGAGAGCCAGTTGAAAATGTGAAGAAGCTAGCTGA 800
Db 821 CTTTTTAAGGTACGAGGATCTCAAAGACGACATCGAGACCAACTTGAAGAGGCTTGCAC 880
QY 801 GTTCAATGGAGTGGGTTTCAAGACGATGAGGAGAAAACAAGGGATTTGTTGATGAGATGT 860
Db 881 TTTCTTAGAGCTTCCTTTCCACGGAAGAAGAGCAACGAAGGGAGTTGTGAAGCTATCCG 940
QY 861 TAACTTTGTAGCTTCGCAATCTGAAGAATCAACAGGTGAACAAAACGATCAAGCTA 920
Db 941 CGAGCTGTAGCTTCGAGAAATCTGAAGAAGTTGGAGGTGAAC---AAGTCAAAACAGTC 997
QY 921 CAATTCGAAAATCGACAACAGCATTTCTTCAGGAAGGTGAGGTGAGAGATTTGGCAAA 980
Db 998 GATCAAGAACTTTGAGAATCGAATCTTGTTCGGAAGGAGAAAGTGAATTTGGGTTAA 1057
QY 981 CTATCT 986
Db 1058 CTATTT 1063

RESULT 8
AAF29177
```

ID AAF29177 standard; DNA; 1077 BP.
 AC AAF29177;
 XX
 DT 09-APR-2001 (first entry)
 XX
 DE Hydroxyjasmonic acid sulfotransferase AtST2a gene.
 XX
 KW Hydroxyjasmonic acid sulfotransferase; AtST2a; flowering time; cabbage;
 XX jasmonate; genetically modified plant; lettuce; sugar cane; carrot;
 KW increase vegetative growth; biomass increase; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200102589-A2.
 XX
 PD 11-JAN-2001.
 XX
 PF 06-JUL-2000; 2000WO-CA00801.
 XX
 PR 06-JUL-1999; 99CA-2274873.
 XX
 PA (VARIN/) VARIN L.
 XX (GIDD/) GIDDA S.
 XX
 PI Varin L, Gidda S;
 XX
 DR WPI; 2001-159272/16.
 DR P-PSDB; AAB49722.
 XX
 XX Methods for modulating flowering in plants, particularly useful for
 PT plants used in the food-processing industry, involves modifying the
 PT endogenous level of compounds of the jasmonate family
 XX
 PS Claim 38; Fig 7; 50pp; English.
 CC
 CC This invention relates to a method for modulating flowering in a plant.
 CC The method comprises modifying the endogenous level of at least one
 CC compound of the jasmonate family in a plant. The methods are used to
 CC produce plants which are genetically modified to flower early or tardily
 CC when compared to a corresponding plant that is not genetically modified,
 CC where the modified plant has an increased (flower early) or lowered
 CC (flower tardily) level of jasmonic acid, or a compound of the jasmonate
 CC family. The method is useful for modulating flowering, particularly for
 CC plants that are used in the food-processing industry and plants with
 CC horticultural value. The method is particularly useful for e.g. delaying
 CC flowering time in crops like lettuce, cabbage, sugar cane or carrots,
 CC which results in increased vegetative growth and biomass. The present
 CC sequence represents the Arabidopsis thaliana AtST2a gene, which encodes
 CC an hydroxyjasmonic acid sulfotransferase protein, which can be used in
 CC the method of the invention.
 XX
 SQ Sequence 1077 BP; 294 A; 257 C; 254 G; 272 T; 0 other;
 Query Match 8.5%; Score 100.8; DB 22; Length 1077;
 Best Local Similarity 48.6%; Pred. No. 1.4e-14;
 Matches 440; Conservative 0; Mismatches 442; Indels 24; Gaps 5;
 QY 96 GCAGAGAGAGAGAGATCCAAATGTACAGAGATATAGAGAGATGTTCTTCACCT 155
 DB 102 GAAAGCGGAGAGAGATCAAGGGCTAAGCTGCGAGTTCACAGAGATGTTGGATTCTCTCC 161
 QY 156 TCCTCGAATGATTATTGGGGGATACCATGAGGTTGTACAGGGGATTTGGCAATGG 215
 DB 162 TAAGGAGAGAGATGAGAACTCGTTACCTTACCTATTCAGAGGTTTGGTCCCAAGC 221
 QY 216 ATATCTTGATCTGATCATGCTTTTCAAGAGATAATTTCAAGGCTCGAGAGACGACAT 275
 DB 222 CAAAGAGATTCAAGCCATCATGCTTTTCCAAAACATTTCCATCCCTCGAAAACGACGT 281
 QY 276 TATCCTTAGACATCTTCGAAAGGCTCGAAGCATGAGCAAGGCACTGACGTTTCCCAT 335
 DB 282 CGTTCTCGCCACCATACCTAAATCCGGTACAACTCGGCTAAAAGCTTTAACTTTCCACCAT 341

QY 336 CCTAACACGAGA-----TGTTAACCCACCATCATCACGACACATCCACATTTTGTCTT 389
 DB |||||
 QY 342 CCTTAACCGTCACCGTTTGTATCCGGTTGCTCGAGTACCAACACCCCTCTTTTTCACITC 401
 DB |||||
 QY 390 CAACCTCATTCGTGTGTTCAAAATTTGGAGTATTTGTACATGGGTAGAGAAAATACGAT 449
 DB |||||
 QY 402 CAACCTCATACCTTGTACCTTTCTCGAGTACAAAGCTTTACGCCAACGAGATGTTCC 461
 DB |||||
 QY 450 GCCAGACCTCGATATGTTGAATCGCCGAGGTTGTTGCCGACACATCCCATCTC 509
 DB |||||
 QY 462 CGATCTCTCGGCTAGCCAGT-----CCAAGAACGTTTCCCAACCTTATACCGTTCGG 515
 DB |||||
 QY 510 TTTGTTGCCGCTCTGTTTGAATCGGGAACAAAATCATATAAAGCCGCAACCG 569
 DB |||||
 QY 516 TTCCCTAAAGGNAACGATCGAGAAACCGGCTCGTGTACTTGTGCGCGNACCC 575
 DB |||||
 QY 570 TAAGAGTACATTTGTTCTTTTGGAAATTTGGCAATCTGATTAAACCCCGA-----CAA 623
 DB |||||
 QY 576 GTTTCACACATTCATCTCTTCTGTCGATTTACACCAACAAACATCAATCCGAGTCAGTGAG 635
 DB |||||
 QY 624 GTTATTGGACCTCGAAAGAGCGTTGATATCTTCCATCGGGAATCTCTTTTGTGGACC 683
 DB |||||
 QY 636 CCCAGTCTTGTAGACCAAGCTTTTGTATCTGTTGCGGGGAGTGATCGGGTTTGGCCC 695
 DB |||||
 QY 684 GGAATGGAATTTCCAAAGCGGAGTTTCAACAATCGCGCTCTACTAATTCA---AACTTGTCT 740
 DB |||||
 QY 696 GTTTTGGGAACACATGTTTGGGATCTCGAGAGAGAGCTTGAAGACCCAGAGAAAGTCTT 755
 DB |||||
 QY 741 ATTGTTGAGTTACGAAGAAATGTTTAGAAGCCAGTTCGAAATGTGAAGAGCTAGTGA 800
 DB |||||
 QY 756 CTTTTTAAGGTACGAGGATCTCAAAAGAGACATCGAGACCAACTTGAAGAGGCTTGCAC 815
 DB |||||
 QY 801 GTTCATGGAGTGGGTTTCACAGAGATGAGGAGAAACAGGGATTTGTAGAGATAGT 860
 DB |||||
 QY 816 TTCTTAGAGCTTCTTTCACCGAAGAGAGAGAACGAAAGGGAGTTGTGAAGGCTATCGC 875
 DB |||||
 QY 861 TAACTTTGTAGCTTCGACAATCTGAAGATCAACAGGTGAACAAAACCGATCAAGCTA 920
 DB |||||
 QY 876 CGAGCTGTGAGCTTCGAGATCTGAAGAGTTGGAGGTGAAC---AAGTCAACAAGTC 932
 DB |||||
 QY 921 CAATTTCGAAATTCGACAAACAGCATTTCTTCAGAAAGGTGAGGTGAGAGATTTGGGCAA 980
 DB |||||
 QY 933 GATCAAGAACTTTGAGAAATCGAATTCGATTCTTGTTCGAAAGGAGAGAGTGAAGTGGGTTAA 992
 DB |||||
 QY 981 CTATCT 986
 DB |||||
 QY 993 CTATTT 998
 DB |||||
 RESULT 9
 AAC48839
 ID AAC48839 standard; DNA; 1270 BP.
 XX
 AC AAC48839;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 58964.
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 22-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

Query Match		8.58;	Score 100.8;	DB 21;	Length 1270;
Best Local Similarity		48.66;	Pred. No. 1.4e-14;		
Matches 440;		Conservative 0;	Mismatches 442;	Indels 24;	Gaps 5;
Pr	05-OCT-1999;	99US-0157753.			
Pr	06-OCT-1999;	99US-0157865.			
Pr	07-OCT-1999;	99US-0158029.			
Pr	08-OCT-1999;	99US-0158232.			
Pr	12-OCT-1999;	99US-0158369.			
Pr	13-OCT-1999;	99US-0159293.			
Pr	13-OCT-1999;	99US-0159295.			
Pr	13-OCT-1999;	99US-0159294.			
Pr	14-OCT-1999;	99US-0159329.			
Pr	14-OCT-1999;	99US-0159330.			
Pr	14-OCT-1999;	99US-0159331.			
Pr	14-OCT-1999;	99US-0159637.			
Pr	14-OCT-1999;	99US-0159638.			
Pr	18-OCT-1999;	99US-0159584.			
Pr	21-OCT-1999;	99US-0160741.			
Pr	21-OCT-1999;	99US-0160767.			
Pr	21-OCT-1999;	99US-0160768.			
Pr	21-OCT-1999;	99US-0160770.			
Pr	21-OCT-1999;	99US-0160814.			
Pr	21-OCT-1999;	99US-0160815.			
Pr	22-OCT-1999;	99US-0160980.			
Pr	22-OCT-1999;	99US-0160981.			
Pr	22-OCT-1999;	99US-0160989.			
Pr	25-OCT-1999;	99US-0161404.			
Pr	25-OCT-1999;	99US-0161405.			
Pr	25-OCT-1999;	99US-0161406.			
Pr	26-OCT-1999;	99US-0161359.			
Pr	26-OCT-1999;	99US-0161360.			
Pr	26-OCT-1999;	99US-0161361.			
Pr	28-OCT-1999;	99US-0161920.			
Pr	28-OCT-1999;	99US-0161992.			
Pr	28-OCT-1999;	99US-0161993.			
Pr	29-OCT-1999;	99US-0162142.			
Qy	96	GCAAGAGAAGGAAGATCCAAATGTCAAGAGATATAGAGAGATGTCTTCACT	155		
Db	165	GAAGCCGAAGAGATGAAGGCTAAGCTGGAGTTCAGAGATGTGTGATCTCTTC	224		
Qy	156	TCCCTCGAATGATTTATTTGGGGGATACCATGAGTTGTACAAGGATTTTGGCAATGG	215		
Db	225	TAAGGAGAGAGATGAGAACTCGTTACCTTACCTATTCCAAGGTTTGGTGCCAAGC	284		
Qy	216	ATATCTTGACCTGGTATCATGCTTTCGAAGTAAATTTCAAGCTCGAGAGAGCAT	275		
Db	285	CAAGAGATCAAGCCCATCATGCTTTCCTCAAAACATTTCCAATCCCTCGAAGACGACGT	344		
Qy	276	TATCCTTTACGACTCTTCCAAAGGCTGGAAGCATGACGAGCACTGACGTTTGGCAT	335		
Db	345	CGTCTCGCCACCATACTTAATCCGCTACAACTGGCTAAAGCTTTAACTTTCACCAT	404		
Qy	336	CCTAACACGAGA-----TGTTAACCAACCATCATCACCGACATCCACTTTTGTCTT	389		
Db	405	CCTTAACCGTCACCGGTTTGTATCGGTTGCTCGAGTACCAACACCCCTCTTTCACTTC	464		
Qy	390	CAACCCCTCATTCGTGCTGTCCTCAAAATTTGGAGTATTTGTATCGGTAGAGAAATACGAT	449		
Db	465	CAACCCCTCATGACCTTGATCTTCTTCGAGTCAAGCTTTACGCCAACGGAGATGTTC	524		
Qy	450	GCCAGACCTCGATATGTTGAATGAATCGCGAGGTTTGTTCGGGACACATCCCATACTC	509		
Db	525	CGATCTCTCGGCTAGCCAGT-----CCAAGAACGTTCCGAACCCACTTACCGTTCGG	578		
Qy	510	TTTGTTCGGCGTCTGTTTGAATCGGGAAACAAATCATCAATATAAGCCGCAACCG	569		
Db	579	TTCCCTTAAGGAAACGATCGAGAAACCCCGTGTGAGGTCGTGTACTTGTGCGCGGAACCC	638		
Qy	570	TAAGAGTACATTTGTGTCTTTTGGAAATTTGGCAATCTGATTAACCCCGA-----CA	623		
Db	639	GTTTGACACATTCATCTCTTCGTGGCATTAACCAACACATCAATCCGAGTCAGTGAG	698		

Qy	624	GTATTGGACCTCGAAAAGAGCGCTTGATATCTTCGATCGGGAATCTCTTTTGTGACC	683		
Db	699	CCAGTCTTGTAGACCAAGCTTTTGTATCTGTATTCCGGGGAGTGATCGGGTTTGGCCC	758		
Qy	684	GGAATGGAATTTCCAAAGCGAGTTCAACCAATCGCGCTCTACTAAATTCA---AACTTGTCT	740		
Db	759	GTTTTGGGAACACATGTTGGGATCTGGAGAGAGAGCTTTGAAGAGACCAAGAGTCTTT	818		
Qy	741	ATTGTTGAGTTACGAAGAAATGTTAGAGAGCCAGTTGAAAATGTGAAGAGCTAGCTGA	800		
Db	819	CTTTTAAAGGTACGAGGATCTCAAGACGACATCGAGACCAACTTGAAGAGGCTTGCAC	878		
Qy	801	GTTTCATGGGATGTTGGGTTCCAGACGATGAGAGAAACAGGATTTGTGTAGATAGT	860		
Db	879	TTTCTTAGAGCTTCTCTTCCCGAAGAGAGAAACGAAAGGGAGTTGTGAAGCTTATCGC	938		
Qy	861	TAACTTTTGTAGCTTCGACAATCTGAAGAAATCAACAGGTGAACAAACCGATCAAGCTA	920		
Db	939	CGAGCTGTGTAGCTTCGAGAACTCTGAAGATTTGGAGGTGAAC---AAGTCAACCAAGTC	995		
Qy	921	CAATTCGAAAATCGACAACAAAGCATTTCTTCAGGAAAGGTGAGGTGAGAGATTGGGCAA	980		
Db	996	GATCAAGAACTTTGAGATCGATTTCTTGTTCGAAAGGAGAAGTGATGATTTGGGTAA	1055		
Qy	981	CTATCT 986			
Db	1056	CTATTT 1061			
RESULT 10					
AAC37555					
ID	AAC37555 standard; DNA; 1347 BP.				
XX	AAC37555;				
XX	DT 17-OCT-2000 (first entry)				
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 17816.				
DE	Hybridisation assay; genetic mapping; gene expression control;				
KW	protein identification; signal transduction pathway;				
KW	metabolic pathway; promoter; termination sequence; ss.				
XX	Arabidopsis thaliana.				
XX	EP1033405-A2.				
XX	06-SEP-2000.				
XX	25-FEB-2000; 2000EP-0301439.				
PF	99US-0121825.				
XX	99US-0123180.				
PR	99US-0123548.				
PR	99US-0125788.				
PR	99US-0126264.				
PR	99US-0126785.				
PR	99US-0127462.				
PR	99US-0128234.				
PR	99US-0128714.				
PR	99US-0129845.				
PR	99US-0130077.				
PR	99US-0130449.				
PR	99US-0130510.				
PR	99US-0130891.				
PR	99US-0131449.				
PR	99US-0132048.				
PR	99US-0132407.				
PR	99US-0132484.				
PR	99US-0132485.				
PR	99US-0132486.				
PR	99US-0132487.				

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140821.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 7.9%; Score 94.4; DB 21; Length 1347;
Best Local Similarity 49.2%; Pred. No. 5.1e-13;
Matches 446; Conservative 0; Mismatches 421; Indels 39; Gaps 6;
QY 118 AAAATGTACAAGAGATATAGAGAGATTGTTTCTTCACTTCCCTCGAATGATTATTGG---174
DB 247 AAAACCCAGAGAGATATCAAGATTTCATCGCTACACTTCCAAGAGCAAGGCTGAGA 306
QY 175 ---GGGATACCATGAGGTTGTACAAGGGATTTTGGCAATGGATATCTTTGACCTGGT 231
DB 307 CCAGATGAGATCTTAACCCAAATACGGTGGACACTGGTGGCAAGATGTCTCTCGAAGGT 366
QY 232 ATCATGGCTTTCGAAGATATTTCAAGGCTCGAGACGACATATCTTACGACTCTT 291
DB 367 CTTTTTCAAGCACTATTTTGAAGCAAGCACTGATTTCTCTCGTCTAGCTAC 426
QY 292 CCAAAGCTCGAAGCATCGAAGGCACTGACGTTTGGCCATCTTAACACGAGATGTT 351
DB 427 CCAAAAACCGTACAACTTGGCTCAAGCACTAATCTACGCAATCGTCAATCGTCTCGT 486
QY 352 AACCAACCATCATCACCGACACATCCACTTTGTTCTTCAACCTCATCTCGTGTGTTCAA 411
DB 487 TACGACGA-----CGCCGCAACCCACTCTCAACGAAACCCCTCACGAGTTTGTCCCT 540
QY 412 AATTGGAGTATTTGTACATGGTGGTGAAGAAATACGATGCCAGACCTCGATATGTTGAAT 471
DB 541 TACGTTGAGATCGACTTTCGGGTTTACCCCA-----CCGTTGATGTTCTTCAAGAC 591
QY 472 GAATCGCGAGTGTGTTGCGGACACATCCCATCTTTGTTGCGGGGCTGTTTGTG 531
DB 592 AGAAGAACCCACTTCTCTACTCATATCCAAACGGTTATTACCGATTCGATGTTG 651
QY 532 AAATCGGGAACAAAATCATCAATATAAGCCGCAACCGTGAAGATACATTTGTGCTTTT 591
DB 652 AACTCTGGTTGTAAGATGGTGTACATATGGAGAGACCCGAAAGATACCTTTCATCTCCATG 711
QY 592 TGGAAATTTGCAATCTGATTAACCCGACA-----AGTTATGGACCTCGAAAGAGC 645
DB 712 TGGAAATTTCTTACCAAGGAGAGTCTCAAGAGGTCAATTAGCGAGTCTTTGAGGACAGC 771
QY 646 GTTGATATCTTCGCATCGGGAATCTCTTTTGTGGACCGGAATGGAATTTTCAAGCCGAG 705
DB 772 TTTGATATGTTTGTAAAGTTTATCTGTGATGTCCTTATCTGGATCATGTTTGGGT 831
QY 706 TTCACCAATGGCGGCTCTACTAATTCAACTTGTCTATTTGT---TGATTAGCAAGAATG 762
DB 832 TATTGGAAGCTTTACCAAGAGAAATCCAGATAGGATTTTGTTCCTTAGTACGAGACCATG 891
QY 763 TTAGAGAGCCAGTTGAAATGTGAAGAGCTAGCTGAGTTTCATGGGATGGGGTTCACA 822
DB 892 AGGCCCAATCTTTGCCCTTTTGTGAAGAGATTGGCTGAGTTTCATGGTTATGGATTCACT 951
QY 823 GACCATGAGGAGAAACAAGGAGATTGTTGATGAGATAGTTAAACTTTGTAGCTTCGACAA 882
DB 952 GATGAGGAAGAGGAGAAATGGTGTCTGAGAAAGTGGTGAAGCTTTGTAGCTTTGAGACG 1011
QY 883 CTGAAGATCAACAGGTGAC-----AAAAACGGATCAAGCTACAAATTCGAAATC 933
DB 1012 TTGAAGAATCTTGAAGCTTAACAAAGGTGATAAAGAAAGAGAGGATCGCTCTGCTTTAT 1071

QY 934 GACAACAAGCATTTCTTTCAGGAAAGGTGAGCTGAGAGATTGGCAAACTATCTTAACGTGC 993
DB 1072 CGAATACGCGGTATTTAGCAAGAAAGGTTGGAGATTGGCTAATTAATTGACTCT 1131
QY 994 GAAATG 999
DB 1132 GAGATG 1137
RESULT 11
ID AAF29178 standard; DNA; 1041 BP.
XX AAF29178;
XX 09-APR-2001 (first entry)
XX Hydroxyjaemomic acid sulfotransferase AtST2b gene.
XX Hydroxyjaemomic acid sulfotransferase; AtST2b; flowering time; cabbage;
KW jasmonate; genetically modified plant; lettuce; sugar cane; carrot;
KW increase vegetative growth; biomass increase; ds.
XX Arabidopsis thaliana.
XX WO200102589-A2.
XX 11-JAN-2001.
XX 06-JUL-2000; 2000WO-CA00801.
XX 06-JUL-1999; 99CA-2274873.
XX (VARI/) VARIN L.
XX (GIDD/) GIDDA S.
XX Varin L, Gidda S;
XX WPI: 2001-159272/16.
XX P-PSDB; AAB49723.
XX Methods for modulating flowering in plants, particularly useful for
XX plants used in the food-processing industry, involves modifying the
XX endogenous level of compounds of the jasmonate family
XX Claim 38; Fig 9; 50pp; English.
XX This invention relates to a method for modulating flowering in a plant.
XX The method comprises modifying the endogenous level of at least one
XX compound of the jasmonate family in a plant. The methods are used to
XX produce plants which are genetically modified to flower early or tardily
XX when compared to a corresponding plant that is not genetically modified,
XX where the modified plant has an increased (flower early) or lowered
XX (flower tardily) level of jasmonic acid, or a compound of the jasmonate
XX family. The method is useful for modulating flowering, particularly for
XX plants that are used in the food-processing industry and plants with
XX horticultural value. The method is particularly useful for e.g. delaying
XX flowering time in crops like lettuce, cabbage, sugar cane or carrots,
XX which results in increased vegetative growth and biomass. The present
XX sequence represents the Arabidopsis thaliana AtST2b gene, which encodes
XX an hydroxyjasmonic acid sulfotransferase protein, which can be used in
XX the method of the invention.
XX Sequence 1041 BP; 287 A; 241 C; 242 G; 271 T; 0 other;
Query Match 7.5%; Score 89.2; DB 22; Length 1041;
Best Local Similarity 47.8%; Pred. No. 8.5e-12;
Matches 438; Conservative 0; Mismatches 448; Indels 30; Gaps 5;
QY 98 AAGAGAGAGGAAAGATTCCAAAATGTACAAGAGATATAGAGAGATTGTTTCTTCACTT 157
DB 62 AAGCCCAAGAGAGAGGCGCTAAGCTACGAGTTCCAGAGAGATTGTGACTCTCTCTTA 121

QY 158 CCTCGAATGATTATTTGGGGGATACCATGAGTTGTCAAGGGATTTTGGCAATGGAT 217
 Db |||||
 QY 122 AGGAGAGAGGACGGAGAAATCGTTACCTTTACTTTATTCGAAGGGTTTCGGTGCCAAAGCTA 181
 Db |||||
 QY 218 ATCTGTGACCTGGTATCATGCTTTTCGAGATATTTCAAGGCTCGAGAGACGACATTA 277
 Db |||||
 QY 182 AGGAGATCAAGCTATCATCGCTTTTCCAAAAACATTTTCAGTCCCTTCCAGACGACGTG 241
 Db |||||
 QY 278 TCCTTACGACTCTTCCAAAGGCTGGAACGACATCGAAGGCACTGACGTTTGCCTATCC 337
 Db |||||
 QY 242 TCCTCGCCACCATACCTAAATCTGCACAACTGGTTAAAGCTTTAACTTTTCCACCATCC 301
 Db |||||
 QY 338 TAAACAGGATGTTAAACCCCATCATCACCGACA-----CATCCACTTTTGTGTCT 388
 Db |||||
 QY 302 TTACCCGTCATCGGTTTGATCCGGTTTCTCATCAAGTTCGACACCCCTTCTCTCAT 361
 Db |||||
 QY 389 TCAACCTCATCTCGTGTGTTCAAAATTTGGAGTATTTGTACATGGGTAGAGAAATACGA 448
 Db |||||
 QY 362 CCAACCTTCAGACTCGTACCTTTCTTCGAGTACAAGCTTTACGCCAACGGAATGTC 421
 Db |||||
 QY 449 TGCAGACCTCGATATGTTGAATGAATCGCCGAGGTTGTTGCGGACACATCCCATCT 508
 Db |||||
 QY 422 CCGATCTCTCGGCTAGCCAGT-----CCAAGAACTTCGCAACCCAGTACCGTTG 475
 Db |||||
 QY 509 CTTTGTTCGCGGCTGTTTGTGAATCGGGAACAAATATCAATATAGCGCGCAACC 568
 Db |||||
 QY 476 GTGCCCTTAAGGATTCGGTTCGAGAATCCAGTGTGAAGGTTGTGTACCTGTGCGGGAACC 535
 Db |||||
 QY 569 GTAAGAGTACATTTGTCTTTTGGAAATTTGGCAATCTGATTAACCCCGA----- 620
 Db |||||
 QY 536 CGTTTGACATTCATCTCATGTGGCATTACATCAACACATCATCTCCGAGTCAAGTGA 595
 Db |||||
 QY 621 -CAAGTTATTGGACCTCGAAAGAGCGTTGATATCTTCGCATCGGGAATCTCCTTTTGTG 679
 Db |||||
 QY 596 GCGCAGTCTTGCTAGACGAAGCTTTTGATCTATATTCCCGGGATTAATCGGATTTG 655
 Db |||||
 QY 680 GACCGGAATGGAATTTCAAGCGGAGTTCCAAATGCGGCGGTCTACTAATTC-----AACT 736
 Db |||||
 QY 656 GCGCGTTTGGGAACACATCTGCGGATCTCGAGAGAGAGCTTGAAGAGGCCAGAGAAAG 715
 Db |||||
 QY 737 TGCTATTGTTGAGTTAGCAAGAAATGTTAGAGAGCCAGTTGAAATGTGAAGAGCTAG 796
 Db |||||
 QY 716 TCTTATTTTAAAGTAGAGAGATCTCAAGAGACATCGAGACCACTTTGAAGAAGCTAG 775
 Db |||||
 QY 797 CTGAGTTTCATGGATGCGGTTTCAAGACATGAGGAGAAACAAAGGATTTGTGATGAGA 856
 Db |||||
 QY 776 CAAGTTTCTTAGGACTTCCTTTTCCCGAAGAGGAAACAAAGGAGTTGTGAAGCTA 835
 Db |||||
 QY 857 TAGTTAACTTTTGTAGCTTCGAAATCTGAAGATCAACAGGTGAACAAAACGGATCAA 916
 Db |||||
 QY 836 TCCTGTATCTGTGTAGCTTTGGAATCTGAAGAAGTTGGAGGTGAACAAGTCAAGCAAT 895
 Db |||||
 QY 917 GCTACAATTCGAAATTCGACAAAGCATTTCTTCAGGAAGGTGAGGTGAGAGATTGGG 976
 Db |||||
 QY 896 TGATCCA---GAATATGAAACCGGTTCTTGTGTTAGGAAGAGAGTGAAGTTGG 952
 Db |||||
 QY 977 CAAACTATCTAAGCTC 992
 Db |||||
 QY 953 TTAACATTTGTCGCC 968
 Db |||||

RESULT 12

AAC32748

ID AAC32748 standard; DNA; 1160 BP.

XX AC

XX AAC32748;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 503.

XX Hybridisation assay; genetic mapping; gene expression control;

protein identification; signal transduction pathway;
 metabolic pathway; promoter; termination sequence; ss.
 Arabidopsis thaliana.
 EP1033405-A2.
 06-SEP-2000.
 25-FEB-2000; 2000EP-0301439.
 25-FEB-1999; 99US-0121825.
 05-MAR-1999; 99US-0123180.
 09-MAR-1999; 99US-0123548.
 23-MAR-1999; 99US-0125788.
 25-MAR-1999; 99US-0126264.
 29-MAR-1999; 99US-0126785.
 01-APR-1999; 99US-0127462.
 06-APR-1999; 99US-0128234.
 08-APR-1999; 99US-0128714.
 16-APR-1999; 99US-0129845.
 19-APR-1999; 99US-0130077.
 21-APR-1999; 99US-0130449.
 23-APR-1999; 99US-0130510.
 28-APR-1999; 99US-0130891.
 30-APR-1999; 99US-0131449.
 30-APR-1999; 99US-0132048.
 30-APR-1999; 99US-0132407.
 04-MAY-1999; 99US-0132484.
 05-MAY-1999; 99US-0132485.
 06-MAY-1999; 99US-0132486.
 06-MAY-1999; 99US-0132487.
 07-MAY-1999; 99US-0132863.
 11-MAY-1999; 99US-0134256.
 14-MAY-1999; 99US-0134218.
 14-MAY-1999; 99US-0134219.
 14-MAY-1999; 99US-0134221.
 18-MAY-1999; 99US-0134370.
 18-MAY-1999; 99US-0134768.
 19-MAY-1999; 99US-0134941.
 20-MAY-1999; 99US-0135124.
 21-MAY-1999; 99US-0135353.
 24-MAY-1999; 99US-0135629.
 25-MAY-1999; 99US-0136021.
 27-MAY-1999; 99US-0136392.
 28-MAY-1999; 99US-0136782.
 01-JUN-1999; 99US-0137222.
 03-JUN-1999; 99US-0137528.
 04-JUN-1999; 99US-0137502.
 07-JUN-1999; 99US-0137724.
 08-JUN-1999; 99US-0138094.
 10-JUN-1999; 99US-0138540.
 10-JUN-1999; 99US-0138847.
 14-JUN-1999; 99US-0139119.
 16-JUN-1999; 99US-0139452.
 16-JUN-1999; 99US-0139453.
 17-JUN-1999; 99US-0139492.
 18-JUN-1999; 99US-0139454.
 18-JUN-1999; 99US-0139455.
 18-JUN-1999; 99US-0139456.
 18-JUN-1999; 99US-0139457.
 18-JUN-1999; 99US-0139458.
 18-JUN-1999; 99US-0139459.
 18-JUN-1999; 99US-0139460.
 18-JUN-1999; 99US-0139461.
 18-JUN-1999; 99US-0139462.
 18-JUN-1999; 99US-0139463.
 18-JUN-1999; 99US-0139750.
 21-JUN-1999; 99US-0139763.
 21-JUN-1999; 99US-0139817.
 22-JUN-1999; 99US-0139899.
 23-JUN-1999; 99US-0140353.
 23-JUN-1999; 99US-0140354.

diseases associated with DNA transcription, e.g. immunological disorders, Werner syndrome, psoriasis, myocardial infarction, solid tumours or cancer -

Claim 1: SEQ ID No 327; 32pp; English.

The invention relates to a nucleic acid, which comprises a segment of the chemically pretreated DNA of genes associated with DNA transcription from one of 346 sequences, and an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical to the chemically pretreated DNA of genes associated with DNA transcription. The set of oligomer probes are useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) in a chemically pretreated genomic DNA. The nucleic acids are useful for diagnosing or treating diseases associated with DNA transcription (particularly with the methylation status), e.g. adenosine deaminase deficiency, viral infection, retroviral infection, Sezary syndrome, haematological disorders, immunological disorders, Werner syndrome, tuberculosis, developmental disorders, psoriasis, Rieger's syndrome, neurological disorders, neurodegenerative disorders, Wardenburg syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial infarction, hypertension, angiogenesis, erythropoiesis, congenital heart disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours or cancer. Sequences ABK28127-ABK28472 represent DNA transcription associated genomic DNA molecules of the invention.

Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from the European Patent Office.

Sequence 10467 BP: 2666 A; 211 C; 2522 G; 5068 T; 0 other;

Query Match	4.0%;	Score 48.2;	DB 24;	Length 10467;
Best Local Similarity	52.8%;	Pred. No. 0.13;		
Matches 104;	Conservative 0;	Mismatches 93;	Indels 0;	Gaps 0;
QY	995	AAATGATTAAAGAACTGGAGACGGCGGAAAAATAAATCAATCAGAGTAAAGACATTTTAT	1054	
Db	4083	AAATAAACCAAAATAATAAAATATTCAAAAAATAATAATACAAAAAATAAATAA	4024	
QY	1055	TATCGTGAATTAAGAATCTTTACATGAACCTCTCGAAATCTTAAATAATTACTGTGAGAAAT	1114	
Db	4023	TATTTTAAAAAATAATTAATTAATAATTTACTTTTAAAAACAAAAATAAATAACAAATAAAAA	3964	
QY	1115	CGAACTAAATATCTCTTTCTTTATTATTCGTATTTCTATTCGTAATAATAATATTTTCATTTTGT	1174	
Db	3963	AAAAATAAATATCTTAAATTACTTATATACTACTATTTATAAAAAAATACTTTTAT	3904	
QY	1175	TAAAAAATAAAAAAATAAAAA	1191	
Db	3903	ATACTAAATAAATAATCACA	3887	

Search completed: June 10, 2003, 17:28:10
Job time : 319 secs

XX	Sequence	7303 BP;	1683 A;	293 C;	1744 G;	3583 T;	0 other;
QY	Query Match	4.1%;	Score 48.4;	DB 24;	Length 7303;		
Db	Best Local Similarity	46.9%;	Prod. No. 0.1;				
	Matches 151;	Conservative	0;	Mismatches 171;	Indels	0;	Gaps
QY	871	AGTTCGCAAACTGTGAAGAAATCAACAGGTGAACAAAAACGGATCAAGCTACAATTCGAAA	930				
Db	1017	AACTAAAAACAAAATAAAAAAAATAAAAAATTCAAAAAAAATTAATAATATAATTCATATA	958				
QY	931	ATCGACAAACAAGCATTTCTTCAGAAAAGGTGAGGTGAGAGATTGGGCAAAACTATCTTAACG	990				
Db	957	AACCGAAAAAAACCTCATTTAAAAAAAACAAACTTCGAACCAAAATTTAAAAAAAATAAAAAAAT	898				
QY	991	TCGGAAATGATTAAAGAAACTGGAGACGGCCGGAANAATAATCAATCAAGTACAGAGTAAAGCAT	1050				
Db	897	TAATAATACAAATAACTCACTCAAAATACTTCTTAACCAACAACAAAAAAAACCAAC	838				
QY	1051	TTATTATCGTGAATAAGAACTTTACATGAAACTTCTGAAATCTTAATAATTAATCTGTGAG	1110				
Db	837	ATTTAACAATAAAATCACAACCTAAAATATACTTTAAACCGTCACATATACATTTTAAACG	778				
QY	1111	AAATCGAACTAAATATCTCTTTGTTTATTAATCGTATTCATTCGTAAATAAATAATTTCAAT	1170				
Db	777	AAAAATAAAAAATACATATAACCACTTTAAAAATTTCTATTTTACTATAAAAAATAATAAAT	718				
QY	1171	TTGCTTAAAAAAAATAAAAAAA	1192				
Db	717	TCGTTAAAAAAAATTTAAAAAAA	696				

RESULT 15	
ABK28453/C	
ID	ABK28453 standard; DNA; 10467 BP.
XX	
XX	ABK28453;
AC	
XX	
DT	
XX	23 - April - 2002 (first entry)
XX	
XX	DNA transcription associated genomic DNA #164.
DE	
XX	
XX	DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
XX	PNA; cytosine methylation state; SNP; retroviral infection; gene; db;
KW	single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW	viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW	immunological disorder; Werner syndrome; developmental disorder;
KW	psoriasis; Rieger's syndrome; neurological disorder; erythroplasia;
KW	neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW	myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW	angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW	polyglutamine disorder; solid tumour.

XX	Unidentified.
OS	
XX	
XX	WO200192565-A2.
XX	
XX	06-DEC-2001.
PD	
PF	
XX	06-APR-2001; 2001WO-EP03973.
XX	
XX	06-APR-2000; 2000DE-1019058.
PR	07-APR-2000; 2000DE-1019173.
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	(EPIG-) EPIGENOMICS AG.
PA	
XX	Olek A, Piepenbrock C, Berlin K;
PI	
XX	WPI; 2002-090046/12.
DR	
XX	New nucleic acids or oligomers, useful for diagnosing or treating
XX	
PT	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 17:18:31 ; Search time 84 Seconds
(without alignments)
4351.889 Million cell updates/sec

Title: US-09-854-122-15
Perfect score: 1192
Sequence: 1 acgcg9gggaataactggaat.....gttaaaaaaaaaaaaaaaaa 1192

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	65.8	5.5	7218	1	US-08-232-463-14
2	42	3.5	6243	2	US-09-056-075-1
3	41.2	3.5	731	1	US-08-451-405A-2
4	41.2	3.5	19124	2	US-08-487-826B-13
C 5	39.2	3.3	3012	2	US-08-475-427-5
C 6	39.2	3.3	3012	2	US-07-842-165-5
C 7	39.2	3.3	9909	4	US-08-961-527-12
8	39	3.3	1218	2	US-08-731-722-4
C 9	38.6	3.2	1428	2	US-08-624-601-7
10	38.4	3.2	1736	3	US-09-182-816-22
C 11	38.4	3.2	1736	3	US-09-182-816-24
12	38.4	3.2	1736	3	US-09-471-528-22
C 13	38.4	3.2	1736	3	US-09-471-528-22
14	38.4	3.2	1736	3	US-09-634-530-22
C 15	38.4	3.2	1736	3	US-09-634-530-24
16	37.8	3.2	562	1	US-08-329-704-3
17	37.8	3.2	562	2	US-08-472-604-3
18	37.8	3.2	562	2	US-08-486-117-3
19	37.8	3.2	562	4	US-08-477-537-3
20	37.8	3.2	768	1	US-08-567-816A-1
21	37.8	3.2	2672	1	US-08-703-947-1
22	37.6	3.2	289	4	US-09-007-005-17
23	37.6	3.2	289	4	US-09-244-796-17
24	37.6	3.2	1132	3	US-08-894-731-3
C 25	37.4	3.1	319	1	US-07-593-657-6
C 26	37.4	3.1	1241	1	US-07-593-657-6
C 27	37.4	3.1	1241	4	US-08-942-012B-3

C 28	37.4	3.1	55827	4	US-09-813-133A-3	Sequence 3, Appli
29	37	3.1	2007	3	US-08-747-221B-36	Sequence 36, Appl
C 30	37	3.1	2007	3	US-08-747-221B-38	Sequence 38, Appl
31	37	3.1	2007	4	US-09-005-051-36	Sequence 36, Appl
C 32	37	3.1	2007	4	US-09-005-051-38	Sequence 38, Appl
33	37	3.1	2550	6	5258287-23	Patent No. 5258287
34	37	3.1	6124	4	US-08-213-419B-3	Sequence 3, Appli
35	36.8	3.1	2435	4	US-09-306-593-1	Sequence 1, Appli
36	36.6	3.1	4539	1	US-08-119-512-1	Sequence 1, Appli
37	36.6	3.1	4539	1	US-08-488-015B-1	Sequence 1, Appli
38	36.6	3.1	4542	3	US-08-814-412-11	Sequence 11, Appl
C 39	36.6	3.1	51952	3	US-08-947-823-1	Sequence 1, Appli
40	36.4	3.1	26664	4	US-09-564-805-28	Sequence 28, Appl
41	36.2	3.0	257	4	US-09-040-984-48	Sequence 48, Appl
42	36.2	3.0	257	4	US-09-123-912-48	Sequence 48, Appl
43	36.2	3.0	257	4	US-09-643-597-48	Sequence 48, Appl
C 44	36.2	3.0	2960	3	US-08-913-842-3	Sequence 3, Appli
C 45	36	3.0	306	4	US-09-122-400B-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match

5.5%; Score 65.8; DB 1; Length 7218;

```

RESULT 2
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,075
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95238
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6243 base pairs

```


Qy	1087	TGAATCTTAATAATTACTGTGAGAAATCGAAATATCTCTTTGTTATTATCGTAT	1146
Db	597	TAATAATCTCTTTTAATAAAATACCTTTTCTACATTATATTTTTTATTATTTTT	656
Qy	1147	TCATTCGTAATAATAATTTTCATTTTGTAAAAAATAAAAAAAAAA	1192
Db	657	TCTTTAATCATTCAAAAATTTTATTTTTTTTTTAAAAAATAAAAAAAAAA	702

RESULT 4

US-08-487-826B-13
 ; Sequence 13, Application US/08487826B
 ; Patent No. 5993827
 ; GENERAL INFORMATION:
 ; APPLICANT: Sim, Kim L.
 ; APPLICANT: Chitnis, Chetan
 ; APPLICANT: Miller, Louis H.
 ; APPLICANT: Peterson, David S.
 ; APPLICANT: Su, Xin-zhaun
 ; APPLICANT: Wellem, Thomas E.
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 ; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe Martens Olson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,826B
 ; FILING DATE: 10-SEP-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israel, Ned
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH121.001CPI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 19124 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; US-08-487-826B-13

Query Match	3.5%	Score 41.2	DB 2	Length 19124
Best Local Similarity	57.9%	Pred. No. 0.3		
Matches 73	Conservative 0	Mismatches 53	Indels 0	Gaps 0
QY	1066	AAGATCTTCATCGAAACTTCTGAAATCTTATAATTAATCTGTGAGAAATCGAACTAAATA	1125	
Db	260	AAATAAATTAATAATAATATACAATAATTTCTGTATTTTATAAAATATAACTTAATTT	319	
QY	1126	TCCTTTTGTATTATTCGTATTTCATTTCGTAATAATAATTTTCATTTTCTGTAAAAA	1185	
Db	320	CTTATTTTATTATTAACCTTATTCCTTTTAAATTTCTTAATTTCTTTTATGCAACAAAAA	379	
QY	1186	AAAAAA	1191	
Db	380	CATAAA	385	

RESULT 5

US-08-475-427-5/c
; Sequence 5, Application US/08475427
; Patent No. 5859340
; GENERAL INFORMATION:
; APPLICANT: DUBOIS, Michel
; APPLICANT: GRISON, Rene
; APPLICANT: LEGUAY, Jean-Jacques
; APPLICANT: PIGNARD, Annie
; APPLICANT: TOPPAN, Alain
; TITLE OF INVENTION: Recombinant gene coding for a protein
; TITLE OF INVENTION: having endochitinase activity or for a precursor thereof.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,427
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/842,165
; FILING DATE: 01-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00607
; FILING DATE: 21-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90 09460
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/POCKET NUMBER: 16781/564/BEDEL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3012 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2384..2462
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2617..2697
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1942..2383, 2463..2616, 2698..3007)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 1942..2166
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: join(2167..2383, 2463..2616, 2698..3007,
; 2698)
; US-08-475-427-5

Query Match 3.3%; Score 39.2; DB 2; Length 3012;

```
Best Local Similarity 52.4%; Pred. No. 0.48; Indels 78; Mismatches 0; Gaps 0;
Matches 86; Conservative 0;

QY 1029 AATGAATCAGAGTAAAGCAATTTATTCGTGAAATAGAAATCTTACATGAAACTTCTG 1088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1006 AATCATACAAATTTAGGAGTATTTACAAAGTACAAATGAAATTTGTTTAAATGAAATTAATA 947

QY 1089 AATCTTAATAATTTACTGTGAGAAATCGAACTAAATATCTCTTTGTTTATTCGTATTC 1148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 946 ACTCAATAAAATTTACTTAAGAGAAATTTGAAATTTATGAATTAATGATTTATATGATTC 887

QY 1149 ATTCGTAATAATAATTTTCATTTTGTAAATGAAATGAAATGAAATGAAATGAAATGAAAT 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 AGACTTCAAGATTTCTTAATACGGATAAGAAATTTTGACAAA 843

RESULT 6
US-07-842-165-5/c
; Sequence 5, Application US/07842165
; Patent No. 5932698
; GENERAL INFORMATION:
; APPLICANT: DUBOIS, Michel
; APPLICANT: GRISON, Rene
; APPLICANT: LEGUAY, Jean-Jacques
; APPLICANT: PIGNARD, Annie
; APPLICANT: TOPPAN, Alain
; TITLE OF INVENTION: Recombinant gene coding for a protein
; TITLE OF INVENTION: having endochitinase activity or for a precursor thereof.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: King Street Station, Suite 500, 1800 Diagonal
; STREET: Road, PO Box 299
; CITY: ALEXANDRIA
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/842.165
; FILING DATE: 19920501
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 16781/348
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3012 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2384..2462
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2617..2697
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1942)..2383, 2463..2616, 2698..3007)
; FEATURE:
```

```
; NAME/KEY: sig_peptide
; LOCATION: 1942..2166
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: join(2167)..2383, 2463..2616, 2698..3007, 2698)
US-07-842-165-5
Query Match 3.3%; Score 39.2; DB 2; Length 3012;
Best Local Similarity 52.4%; Pred. No. 0.48;
Matches 86; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1029 AATGAATCAGAGTAAAGCAATTTATTCGTGAAATAGAAATCTTACATGAAACTTCTG 1088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1006 AATCATACAAATTTAGGAGTATTTACAAAGTACAAATGAAATTTGTTTAAATGAAATTAATA 947

QY 1089 AATCTTAATAATTTACTGTGAGAAATCGAACTAAATATCTCTTTGTTTATTCGTATTC 1148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 946 ACTCAATAAAATTTACTTAAGAGAAATTTGAAATTTATGAATTAATGATTTATATGATTC 887

QY 1149 ATTCGTAATAATAATTTTCATTTTGTAAATGAAATGAAATGAAATGAAATGAAATGAAAT 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 AGACTTCAAGATTTCTTAATACGGATAAGAAATTTTGACAAA 843

RESULT 7
US-08-961-527-12/c
; Sequence 12, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-12

Query Match 3.3%; Score 39.2; DB 4; Length 9909;
Best Local Similarity 53.7%; Pred. No. 0.77;
Matches 80; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1013 AGACGGCCGGAATAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1155 ATAGGGCTTACATATATAGAGAGGTTGCTGTCATGACATAGATATTACTGGAAGAAAGTTG 1096
```



```

; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-182-816-22

Query Match      3.2%; Score 38.4; DB 3; Length 1736;
Best Local Similarity 46.9%; Pred. No. 0.62;
Matches 120; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 937 AACAGCATTCTTCTCAGGAAGGTGAGGTGAGAGATTGGCGAACTATCTAAGCTCGGAA 996
DB 1453 AAGAACCACCTTCTTCTAGCAGAAGATATCTTTAGTCTGTCAAGAAATTTATCGACCACC 1512
QY 997 ATGATTAAGAACTGGAGACGGCGGAAATAAATGAATCAGAGTAAAGCATTTATTA 1056
DB 1513 ATTCCAAAAGACAGCAAAACCAAGAGAATAGAGATCTCTGAGAACTTGTAAATAATT 1572
QY 1057 TCGTGAATAAGAATCTTTACATGAACTTCTGAAATCTTAAATTAATTAATCTGAGAAATCG 1116
DB 1573 ATTTGATTAATAATAATGTTTAAATAAATGAATTAATCTGAAATAAAGCATATGG 1632
QY 1117 AACTAAATATCTCTTTGTTTATTCATTCGTAATAATAATTTTCATTTTGTTA 1176
DB 1633 ATTTTATTCAACTTGTCAAAATATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1692
QY 1177 AAAAAAATAAAAAA 1192
DB 1693 AAAAAAATAAAAAA 1708

RESULT 11
US-09-182-816-24/C
; Sequence 24, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; * APPLICANT: Wisniewski, Nancy
; * APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-182-816-24

Query Match      3.2%; Score 38.4; DB 3; Length 1736;
Best Local Similarity 46.9%; Pred. No. 0.62;
Matches 120; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 937 AACAGCATTCTTCTCAGGAAGGTGAGGTGAGAGATTGGCGAACTATCTAAGCTCGGAA 996
DB 1453 AAGAACCACCTTCTTCTAGCAGAAGATATCTTTAGTCTGTCAAGAAATTTATCGACCACC 1512
QY 997 ATGATTAAGAACTGGAGACGGCGGAAATAAATGAATCAGAGTAAAGCATTTATTA 1056
DB 1513 ATTCCAAAAGACAGCAAAACCAAGAGAATAGAGATCTCTGAGAACTTGTAAATAATT 1572
QY 1057 TCGTGAATAAGAATCTTTACATGAACTTCTGAAATCTTAAATTAATTAATCTGAGAAATCG 1116
DB 1573 ATTTGATTAATAATAATGTTTAAATAAATGAATTAATCTGAAATAAAGCATATGG 1632
QY 1117 AACTAAATATCTCTTTGTTTATTCATTCGTAATAATAATTTTCATTTTGTTA 1176
DB 1633 ATTTTATTCAACTTGTCAAAATATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1692
QY 1177 AAAAAAATAAAAAA 1192
DB 1693 AAAAAAATAAAAAA 1708

RESULT 12
US-09-471-528-22
; Sequence 22, Application US/09471528
; Patent No. 6153397
; GENERAL INFORMATION:
; * APPLICANT: Wisniewski, Nancy
; * APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/471,528
; CURRENT FILING DATE: 1999-12-27
; EARLIER APPLICATION NUMBER: 09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-471-528-22

Query Match      3.2%; Score 38.4; DB 3; Length 1736;
Best Local Similarity 46.9%; Pred. No. 0.62;
Matches 120; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 937 AACAGCATTCTTCTCAGGAAGGTGAGGTGAGAGATTGGCGAACTATCTAAGCTCGGAA 996
DB 1453 AAGAACCACCTTCTTCTAGCAGAAGATATCTTTAGTCTGTCAAGAAATTTATCGACCACC 1512
QY 997 ATGATTAAGAACTGGAGACGGCGGAAATAAATGAATCAGAGTAAAGCATTTATTA 1056
DB 1513 ATTCCAAAAGACAGCAAAACCAAGAGAATAGAGATCTCTGAGAACTTGTAAATAATT 1572
QY 1057 TCGTGAATAAGAATCTTTACATGAACTTCTGAAATCTTAAATTAATTAATCTGAGAAATCG 1116
DB 1573 ATTTGATTAATAATAATGTTTAAATAAATGAATTAATCTGAAATAAAGCATATGG 1632
QY 1117 AACTAAATATCTCTTTGTTTATTCATTCGTAATAATAATTTTCATTTTGTTA 1176
DB 1633 ATTTTATTCAACTTGTCAAAATATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1692
QY 1177 AAAAAAATAAAAAA 1192
DB 1693 AAAAAAATAAAAAA 1708

RESULT 13
```

```

; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-634-530-22

Query Match      3.2%; Score 38.4; DB 4; Length 1736;
Best Local Similarity 46.9%; Pred. No. 0.62;
Matches 120; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 937 AACAAAGCATTTCTTCAGGAAAGGTGAGGTGAGAGATTGGGCAAACTATCTTAACGTGCGAA 996
DB 1453 AAGAACCACTTCCTCTAGCAGAGATATCTTTAGTGTCTCAAGAAATTTTATCGACCACC 1512
QY 997 ATGATTAAAGAACTGGAGACGGCCGGGAAAAATAAATGAATCAGAGTAAAGCATTTTATTA 1056
DB 1513 ATTCAAAAAAGACAGCAAAAACCAAGAGAAATAGAGATCTCTGAGAACTTGTATAAAT 1572
QY 1057 TCGTGAATAAGAACTTTACATGAACTCTGAAATCTTTAATAATTAATTAATGTGAGAAATCG 1116
DB 1573 ATTTGTGATAATAATATAATGTTAAAAATAAATGTAATTAATCTGTGAAATAAACGATATGG 1632
QY 1117 AACTAAATATCTCTTCTGTTTATTATTCGTATTCATTCGTAATAATAATTAATTTCTATTTGTTA 1176
DB 1633 ATTTTATTTCAAACTTGTCAATATATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1692
QY 1177 AAAAAAATAAAAAA 1192
DB 1693 AAAAAAATAAAAAA 1708

RESULT 15
US-09-634-530-24/c
; Sequence 24, Application US/09634530
; Patent No. 6290958
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-Cl-1
; CURRENT APPLICATION NUMBER: US/09/634,530
; CURRENT FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 09/471,528
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 09/182,816
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 08/989,510
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-634-530-24

Query Match      3.2%; Score 38.4; DB 4; Length 1736;
Best Local Similarity 46.9%; Pred. No. 0.62;
Matches 120; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 937 AACAAAGCATTTCTTCAGGAAAGGTGAGGTGAGAGATTGGGCAAACTATCTTAACGTGCGAA 996
DB 284 AAGAACCACTTCCTCTAGCAGAGATATCTTTAGTGTCTCAAGAAATTTTATCGACCACC 225
QY 997 ATGATTAAAGAACTGGAGACGGCCGGGAAAAATAAATGAATCAGAGTAAAGCATTTTATTA 1056
DB 224 ATTCAAAAAAGACAGCAAAAACCAAGAGAAATAGAGATCTCTGAGAACTTGTATAAAT 165
QY 1057 TCGTGAATAAGAACTTTACATGAACTTCTGAAATCTTTAATAATTAATGTGAGAAATCG 1116
DB 164 ATTTGTGATAATAATATAATGTTAAAAATAAATGTAATTAATCTGTGAAATAAACGATATGG 105

```

Qy	1117	AAC TAA ATA TC TCT TTG TTT AT A TA TCG TA TTC AT TCG TA TAA ATA AT TTT CAT TTT GTT A	1176
Db	104	AT TTT ATT T CAA ACT TGT CAA AT TA TAA AAAA AAAAA AAAAA AAAAA AAAAA AAAAA AAAAA	45

Qy	1177	AAAAA 	1192
Db	44	AAAAA 	29

Search completed: June 10, 2003, 18:22:11
Job time : 86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 17:28:17 ; Search time 194 Seconds
(without alignments)
8599.219 Million cell updates/sec

Title: US-09-854-122-15
Perfect score: 1192
Sequence: 1 acgcgggaataactggaat.....gttaaaaaaaaaaaaaa 1192

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_5/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PC1_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PC1US_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1192	100.0	1192	10	US-09-854-122-15
2	127.2	10.7	996	9	US-09-938-842A-1027
3	110.6	9.3	981	9	US-09-938-842A-2049
C 4	74.6	6.3	533	10	US-09-924-035A-540
C 5	55.4	4.6	344	10	US-09-770-791-792
6	46	3.9	461	9	US-09-918-995-15443
C 7	45.4	3.8	743	10	US-09-770-149-48
8	43.2	3.6	525	9	US-10-198-846-1483
9	42.8	3.6	1047	9	US-10-002-344A-2
10	42.6	3.6	1797	9	US-09-974-879-115
11	42.6	3.6	1797	9	US-09-305-736-1172
C 12	42.4	3.6	428	10	US-09-969-373-1172
C 13	42.2	3.5	9515	9	US-10-239-676-159
C 14	41.6	3.5	576	10	US-09-864-761-15424
C 15	41.4	3.5	1960	9	US-09-938-842A-4667
C 16	41.4	3.5	9539	9	US-10-239-676-52
17	41.2	3.5	393	9	US-09-918-995-36329
C 18	41.2	3.5	1757	9	US-09-925-299-39
C 19	41.2	3.5	1757	10	US-09-925-299-39

C 20	41	3.4	23603	9	US-09-860-670-264	Sequence 264, App
C 21	41	3.4	23613	9	US-09-860-670-258	Sequence 258, App
C 22	41	3.4	24218	9	US-09-860-670-263	Sequence 263, App
C 23	40.8	3.4	1380	10	US-09-925-301-318	Sequence 318, App
C 24	40.8	3.4	17421	9	US-10-239-676-53	Sequence 53, Appl
C 25	40.6	3.4	1966	9	US-09-938-842A-3199	Sequence 3199, Ap
C 26	40.4	3.4	344	10	US-09-960-352-1036	Sequence 1036, Ap
C 27	40.4	3.4	458	10	US-09-960-352-14274	Sequence 14274, A
C 28	40.4	3.4	6282	9	US-10-239-676-127	Sequence 127, App
C 29	40.4	3.4	640681	10	US-09-790-988-1	Sequence 1, Appli
C 30	40	3.4	12405	9	US-10-239-676-35	Sequence 35, Appl
C 31	39.8	3.3	6306	9	US-10-239-676-224	Sequence 224, App
C 32	39.8	3.3	17419	9	US-10-239-676-99	Sequence 99, Appl
C 33	39.8	3.3	640681	10	US-09-790-988-1	Sequence 1, Appli
C 34	39.6	3.3	297	10	US-09-969-373-549	Sequence 549, App
C 35	39.6	3.3	12968	9	US-10-239-676-202	Sequence 202, App
C 36	39.4	3.3	2000	9	US-09-938-842A-4722	Sequence 4722, Ap
C 37	39.4	3.3	5689	9	US-10-239-676-90	Sequence 90, Appl
C 38	39.4	3.3	17421	9	US-10-239-676-54	Sequence 54, Appl
C 39	39.4	3.3	127197	9	US-09-754-853A-1	Sequence 1, Appli
C 40	39.4	3.3	1830121	9	US-10-329-960-1	Sequence 1, Appli
C 41	39.2	3.3	433	9	US-09-918-995-4250	Sequence 4250, Ap
C 42	39.2	3.3	499	9	US-10-040-739-1278	Sequence 1278, Ap
C 43	39.2	3.3	634	9	US-10-066-543-345	Sequence 345, App
C 44	39.2	3.3	671	9	US-10-066-543-464	Sequence 464, App
C 45	39.2	3.3	11036	9	US-10-239-676-118	Sequence 118, App

ALIGNMENTS

RESULT 1
US-09-854-122-15
; Sequence 15, Application US/09854122
; Patent No. US20020016980A1
; GENERAL INFORMATION:
; APPLICANT: ALBERTE, RANDALL S.
; APPLICANT: SMITH, ROBERT
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA
; FILE REFERENCE: PHA-007.01
; CURRENT APPLICATION NUMBER: US/09/854,122
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/202,529
; PRIOR FILING DATE: 2000-05-10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 15
; TYPE: DNA
; ORGANISM: Zostera marina
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1..33, 37..1041, 1045..1059, 1063..1077,
; LOCATION: 1081..1086, 1090..1119, 1123..1191)
US-09-854-122-15

Query Match 100.0%; Score 1192; DB 10; Length 1192;
Best Local Similarity 100.0%; Pred. No. 5.4e-286;
Matches 1192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ACGCGGGGAATACCTGGAATCGCTGTGCTAGCTACCACTGATAATGCTGGAAT	60
DB	1	ACGCGGGGAATACCTGGAATCGCTGTGCTAGCTACCACTGATAATGCTGGAAT	60
QY	61	TTAGCTTTGGAGAAATGTTTCGGATCCAGATGACGAGAGAGAGAAATCCAAA	120
DB	61	TTAGCTTTGGAGAAATGTTTCGGATCCAGATGACGAGAGAGAGAAATCCAAA	120
QY	121	ATGTACAAGAGATATAGAGAGATTTCTTTCCTCCGAAATGATTATTTGGGGGAT	180
DB	121	ATGTACAAGAGATATAGAGAGATTTCTTTCCTCCGAAATGATTATTTGGGGGAT	180
QY	181	ACCATGAGTTGTACAAGGATTTTGGCAATGGGATATCTTGTGATCATGCT	240

```
Db 181 ACCATGAGTTGTACAAGGGATTTTGGCAAATGGGATATCTTGCTACCTGGTATCATGGCT 240
Qy 241 TTGGAAGATAATTTCAAGGCTCAGAGACGGACATTTATCTCTACGACTCTTCCAAAGGCT 300
Db 241 TTGGAAGATAATTTCAAGGCTCAGAGACGGACATTTATCTCTACGACTCTTCCAAAGGCT 300
Qy 301 GGAACGACATGGAAGGACGACGTTGGCCATCTTACACGAGATGTTTACCAACCCA 360
Db 301 GGAACGACATGGAAGGACGACGTTGGCCATCTTACACGAGATGTTTACCAACCCA 360
Qy 361 TCATCACCAGACATCCACATTTTGTCTTCAACCCCTCAATTCGTGTGTTTCAAAATTTGGAG 420
Db 361 TCATCACCAGACATCCACATTTTGTCTTCAACCCCTCAATTCGTGTGTTTCAAAATTTGGAG 420
Qy 421 TATTTGTACATGGGTAGAGAAATACGATGCCAGACCTCGATATGTTGAATGAATGCCCG 480
Db 421 TATTTGTACATGGGTAGAGAAATACGATGCCAGACCTCGATATGTTGAATGAATGCCCG 480
Qy 481 AGGTGTTTCCGCGACATCCCATCTCTTGTGTCGGCGCTGTTTGTGAATCGGGA 540
Db 481 AGGTGTTTCCGCGACATCCCATCTCTTGTGTCGGCGCTGTTTGTGAATCGGGA 540
Qy 541 ACAAATATCATATTAAGCCGCAAGTTATGGAAGGACGTTGATATCTTCGCA 600
Db 541 ACAAATATCATATTAAGCCGCAAGTTATGGAAGGACGTTGATATCTTCGCA 600
Qy 601 GGCATCTGATTAACCCGCAAGTTATGGAAGGACGTTGATATCTTCGCA 660
Db 601 GGCATCTGATTAACCCGCAAGTTATGGAAGGACGTTGATATCTTCGCA 660
Qy 661 TCGGGATCTCTTTTGTGACCGGAATGGAATTTCCAAAGGACGTTGATATCTTCGCGG 720
Db 661 TCGGGATCTCTTTTGTGACCGGAATGGAATTTCCAAAGGACGTTGATATCTTCGCGG 720
Qy 721 TCTACTAATTCAAACTTGTCTTGTGAGTTTACGAAGAAATGTTTACGAAGCCAGTTGAA 780
Db 721 TCTACTAATTCAAACTTGTCTTGTGAGTTTACGAAGAAATGTTTACGAAGCCAGTTGAA 780
Qy 781 AATGTGAAGAGCTAGCTGAGTTTATGAGTTTTCGACAAATCTGAAGAAATGTTTACGAAGCCAGTTGAA 840
Db 781 AATGTGAAGAGCTAGCTGAGTTTATGAGTTTTCGACAAATCTGAAGAAATGTTTACGAAGCCAGTTGAA 840
Qy 841 GGGATTTGTGATGAGATGTTTAACTTTTGTAGCTTTCGACAAATCTGAAGAAATGTTTACGAAGCCAGTTG 900
Db 841 GGGATTTGTGATGAGATGTTTAACTTTTGTAGCTTTCGACAAATCTGAAGAAATGTTTACGAAGCCAGTTG 900
Qy 901 AACAAAAACGGATCAAGCTACAATTCGAAAAATCGAACCAAGCAATTTCTTCAGGAAAGGT 960
Db 901 AACAAAAACGGATCAAGCTACAATTCGAAAAATCGAACCAAGCAATTTCTTCAGGAAAGGT 960
Qy 961 GAGGTGAGAGATTTGGGCAACTATCTAACGTCGGAATGATTAGAACTGGAGACGGCC 1020
Db 961 GAGGTGAGAGATTTGGGCAACTATCTAACGTCGGAATGATTAGAACTGGAGACGGCC 1020
Qy 1021 GGAATAATTAATGAATCAGAGTAAAGCAATTTATTCGTGAAATTAAGAAATCTTACATGA 1080
Db 1021 GGAATAATTAATGAATCAGAGTAAAGCAATTTATTCGTGAAATTAAGAAATCTTACATGA 1080
Qy 1081 AACTTCTGAAATCTTAATAATTAATCTGTGAGAAATCGAACTAAATATCTCTTTGTTTATTA 1140
Db 1081 AACTTCTGAAATCTTAATAATTAATCTGTGAGAAATCGAACTAAATATCTCTTTGTTTATTA 1140
Qy 1141 TCGTATTTCGTAATAATAATTAATTTGTTTGTAAAAAATAAAAAA 1192
Db 1141 TCGTATTTCGTAATAATAATTAATTTGTTTGTAAAAAATAAAAAA 1192
```

RESULT 2

US-09-938-842A-1027
; Sequence 1027, Application US/0993842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:

APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SAME, AND METHODS OF USE
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1027
LENGTH: 996
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1027

Query Match 10.7%; Score 127.2; DB 9; Length 996;

Best Local Similarity 50.8%; Pred. No. 1.3e-21;
Matches 468; Conservative 0; Mismatches 433; Indels 21; Gaps 6;

```
Qy 101 AGAAGGAAGAAGATTCCAAATGTACAAGATATAGAGAGATTTGTTCTTCACTTCCCT 160
Db 38 ACATGAAGACGACACGTTAGTCAAGAACCCAGAACTTGATCACTTCTTACCTTCAG 97
Qy 161 CGAATGATTTATGGGGGATACCATGAGTTGTACAAGGATTTTGGCAATGGGATATC 220
Db 98 ACAAGATTTTCATGGTTATGGTCTCTACAACCTACAAGGTTTGTGGTACTATCCAACA 157
Qy 221 TTGTACTGGTATCATGGCTTTTCAAGATAATTTCAAGGCTCGAGACGGACATATCC 280
Db 158 CACTCAAGCCGTTCTTTGACGTCCTCAAAACACCTTCAAGCCACGAGATCTGATATATCC 217
Qy 281 TTACGACTCTTCCAAAGGCTGGAACGACATGACGAAGGCACTGACGTTTGCCATCCTAA 340
Db 218 TCGCTTCTTGGCCAAAGGTGGNACCCTTGGCTCAATCCCTAAATTTTCGCTGTTGATC 277
Qy 341 CACGAGATGTTTAAACCCCATCATCACCGACATCCATTTTGTGTTCTTCAACCCTCAT 400
Db 278 ATAGAAAGATACCCGGAACCCCTCAAAACACATCCTTTGTCTTCAAAAACCCCTCAT 337
Qy 401 CGTGTCTTCAAATTTTGGAGTATTTGTACATCGGTAGAGAAATACGATGCCAGACCTCG 460
Db 338 ACCTTGTCCCATTTCTTTGAGGTTGAGTTTATACGCTA---ATAGCCAAATTTCCGATCTCG 394
Qy 461 ATATGTTGAATGAATCGCGAGGTTGTTTGGCGGACACATCCCATACTCTTTTGTGCGG 520
Db 395 CAAAGT---ATCTTCTCTCTATGATCTTTTACACACATGCACTTACAAGCATTTGCGTG 451
Qy 521 CGTCTGTTTGAATCGGGAACAAAATCATCAATTAAGCCCAACCGTAAGAGTACAT 580
Db 452 AAGCCACCAACAAAGCTTG---CAAAACCGTATATGTGTAGAGGTATCAAAAGATACGT 508
Qy 581 TTGTGTCTTTTGGAAATTTGGCAATCTGATTAACCCCGACAAAGTT-----ATTGACC 634
Db 509 TTGTCTCCGCTGGCATTTAGAAACATGTTTGCATCGCAAGATGGATCAAGCCACTT 568
Qy 635 TCGAAAGAGCGTTGATATCTTCGCATCGGGAATCTCTTTTGTGACCGGAATGGAAT 694
Db 569 TTGAGCTCATGTTTGTGCTTATTTAGAGAGGTTCTCTTATATGACCTTATTTGGGAA 628
Qy 695 TCGAAGCGGAGTTTCAACAATGCGGCGTCT---ACTAATTCAAACTTGTCTTGTGAGTT 751
Db 629 ATGTTATGAGCTATTGGAAAGGAGGCTTGAAGCAAGGAGAAATGTTCTTTTCATGAAGT 689
Qy 752 ACGAAGAAATTTAGAGAACCCAGTTGAAATGTGAAGAGCTAGCTGAGTTTATGGGAT 811
Db 689 ACGAAGAGATAATTTGAGAGCCCTCGTGTTCAGTCAAGACATCGCGGAGTTCTTGGAA 748
```


QY 857 TAGTTAACTTTCTAGCTTCGACAAATCTGAAGAATCAACAGGTGAACAAAAACGGATCAA 916
Db 363 TCTTGGAACTTTTCTTCTTAAGAAACCTTAGCGTTTGAGATCAACAAAAACAGG---AA 307
QY 917 GCTACAATTCGAAATCGACAAACAAAGCATTTCTTCAGAAAGGTGAGGTGAGAGATTGGG 976
Db 306 GCTTGTGCGAAGGAGTAAGTTTCAAGAGTTTTTTTCGTAAGGGGAAGTTGGTGATTGGA 247
QY 977 CAACATCTAATCGTGGAAATGATTAAAGAACTGGAGACG 1017
Db 246 AGAGTTATATGACTCTCGTAATCGAAACAAACAAATCGACATG 206

RESULT 5

US-09-770-791-792/c
; Sequence 792, Application US/09770791
; Patent No. US20020062014A1

; GENERAL INFORMATION:
; APPLICANT: Gorchach, Jörn

; APPLICANT: An, Yong-Qiang

; APPLICANT: Hamilton, Carol M.

; APPLICANT: Price, Jennifer L.

; APPLICANT: Raines, Tracy M.

; APPLICANT: Yu, Yang

; APPLICANT: Rameaka, Joshua G.

; APPLICANT: Page, Amy

; APPLICANT: Matthew, Abraham V.

; APPLICANT: Ledford, Brooke L.

; APPLICANT: Woessner, Jeffrey P.

; APPLICANT: Haas, William David

; APPLICANT: Garcia, Carlos A.

; APPLICANT: Krieker, Maja

; APPLICANT: Slader, Ted

; APPLICANT: Davis, Keith R.

; APPLICANT: Allen, Keith

; APPLICANT: Hoffman, Neil

; APPLICANT: Hurlan, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; FILE REFERENCE: 2029 (PAPA-018PRV)

; CURRENT APPLICATION NUMBER: US/09/770,791

; PRIOR FILING DATE: 2001-01-26

; PRIOR FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 999

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 792

; LENGTH: 344

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-770-791-792

Query Match 4.6%; Score 55.4; DB 10; Length 344;
Best Local Similarity 58.2%; Pred. No. 0.0005;
Matches 117; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

QY 817 TTCACACGATGAGGAGAAACAGGATTTGTTGATGAGATAGTTAACTTTGTTAGCTTC 876
Db 340 TTTACTAAGGAAGAAGAGAGCGGATTCGGTGGATGAGATTATCGATCTTTGTTCTTA 281
QY 877 GACAAATCGAAGAATCAACAGGTGAACAAAAACGATCAAGCTCAAAATTCGAAATCGAC 936
Db 280 CGTAATCTGACGAGTTGGAGATCAATAGACCGG---AAATTTGAATCTGGTAGAGAA 224
QY 937 AACAGCATTTCTTTACGAAAGGTGAGGTGAGAGATTGGGCAAACTATCTAACCTCGGAA 996
Db 223 AACAAATGTTTTTCCGTAAGGAGAGAGTTGTTGATTTGGAAGAACTATTTGACTCTCTGA 164
QY 997 ATGATTAGAACTGGAGACG 1017
Db 163 ATGGAGAACAAATCGACATG 143

RESULT 6

US-09-918-995-15443

; Sequence 15443, Application US/09918995
; Publication No. US20030073623A1

; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 15443

; LENGTH: 461

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: misc feature

; LOCATION: (1)....(461)

; OTHER INFORMATION: n = A,T,C or G

US-09-918-995-15443

Query Match 3.9%; Score 46; DB 9; Length 461;
Best Local Similarity 51.3%; Pred. No. 0.13;

Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 995 AAATGATTAGAACTGGAGACGCGGAAAAATAATGAATCAGAGTAAAGCATTTAT 1054

Db 255 AA 314

QY 1055 TATCGTGAATTAAGATCTTACATGAACCTCTGAAATCTTAATTAATTTACTGTGAGAAAT 1114

Db 315 AA 374

QY 1115 CGAACTAAATATCTCTTTCTTTTATTTATTCGTATTCATTCGTAATAATAATTTTCATTTTGT 1174

Db 375 TAAATTAATANNNTNTTTTTTTTTNTNTNTTAAATAATAATAATAATATATTTAATT 434

QY 1175 TAAAAAAA 1183

Db 435 AAATAAAA 443

RESULT 7

US-09-770-149-48/c

; Sequence 48, Application US/09770149
; Patent No. US20020059663A1

; GENERAL INFORMATION:
; APPLICANT: Gorchach, Jörn

; APPLICANT: An, Yong-Qiang

; APPLICANT: Hamilton, Carol M.

; APPLICANT: Price, Jennifer L.

; APPLICANT: Raines, Tracy M.

; APPLICANT: Yu, Yang

; APPLICANT: Rameaka, Joshua G.

; APPLICANT: Page, Amy

; APPLICANT: Matthew, Abraham V.

; APPLICANT: Ledford, Brooke L.

; APPLICANT: Woessner, Jeffrey P.

; APPLICANT: Haas, William David

; APPLICANT: Garcia, Carlos A.

; APPLICANT: Krieker, Maja

; APPLICANT: Slader, Ted

; APPLICANT: Davis, Keith R.

; APPLICANT: Allen, Keith

; APPLICANT: Hoffman, Neil

; APPLICANT: Hurlan, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; thaliana

```

; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,506
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 743
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-770-149-48

Query Match      3.8%; Score 45.4; DB 10; Length 743;
Best Local Similarity 59.8%; Pred. No. 0.23;
Matches 76; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 1066 AAGAACTCTTACATGAAACTTCTGAAATCTTAATAATTAATCTGTGAGAAATCGAACTAATA 1125
Db 154 ATGTGCTTACAGTACAATTTCTGTAGAACTCCAGATATTTTATGAAACAGATCTTTT 95
Qy 1126 TCTCTTTGTTTATTCATCGTATTCATTCGTAATAATAATTTTCATTTTGTAAAAA 1185
Db 94 TCTCGTTTATACTATCATTTTCATTAATAATAATAATAATAATAATAATAATAATA 35
Qy 1186 AAAAAA 1192
Db 34 AAAAAA 28

RESULT 8
US-10-198-846-1483
; Sequence 1483, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1483
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 46, 52, 75, 78, 79, 80, 85, 91, 100, 102, 107, 109, 114,
; LOCATION: 123, 127, 136, 142, 149, 156, 159, 160, 162, 168, 169, 173,
; LOCATION: 177, 195, 203, 207, 210, 211, 213, 219, 220, 224, 230, 251,
; LOCATION: 261, 268, 277, 309, 327, 333, 337, 339, 355, 359, 364
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 365, 380, 389, 399, 404, 408, 413, 416, 429, 434, 435, 438,
; LOCATION: 440, 458, 473, 479, 490, 511, 519, 520, 522, 525
; OTHER INFORMATION: n = A,T,C or G
; US-10-198-846-1483

Query Match      3.6%; Score 43.2; DB 9; Length 525;
Best Local Similarity 40.5%; Pred. No. 0.67;
Matches 155; Conservative 0; Mismatches 227; Indels 1; Gaps 1;

Qy 811 TGTGGGTTACAGACGATAGGAGAACAAAGGGATGTTTCATGAGATAGTAACTTTGT 870
```

```

Db 70 TTTTNTTNNAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 129
Qy 871 AGCTTCGACAACTCTGAAGAAATCAACAGGTGGAACAAAAACGATCAAGCTACAATTCGAAA 930
Db 130 ACAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 189
Qy 931 ATCGAC-AAACAAGCATTTCTTCAGGAAAGGTGAGGTGAGAGATTGGGCAAACTATCTAAC 989
Db 190 AAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 249
Qy 990 GTCGGAATGATTAAGAAACTGAGACGCGCGGAAAAAATAAATGAATCAGAGTAAAGCA 1049
Db 250 ANAGAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 309
Qy 1050 TTTTATTCCTGTAATAAGAAATCTTACATGAAACTTCTGAAATCTTAAATTAATTAATCTGTGA 1109
Db 310 ATAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 369
Qy 1110 GAAATCGAACTAAATATCTCTTTTGTATTATTCGTATTCGTATTAATAATAATTTTCAT 1169
Db 370 AAAAAATAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 429
Qy 1170 TTTGTTTAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 1192
Db 430 AAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 452

RESULT 9
US-10-002-344A-2
; Sequence 2, Application US/10002344A
; Patent No. US20020172959A1
; GENERAL INFORMATION:
; APPLICANT: Recipon, Herve
; APPLICANT: Sun, Yongming
; APPLICANT: Chen, Sei-yu
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prote
; FILE REFERENCE: DEX-0241
; CURRENT APPLICATION NUMBER: US/10/002,344A
; CURRENT FILING DATE: 2002-03-15
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (125)-(186)
; OTHER INFORMATION: n = a, c, g or t
; US-10-002-344A-2

Query Match      3.6%; Score 42.8; DB 9; Length 1047;
Best Local Similarity 58.0%; Pred. No. 1.2;
Matches 94; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

Qy 1021 GGAATAATAATGAATCAGAGTAAAGCAATTTATTATTCGTGAATAAGAAATCTTACATGA 1080
Db 838 GAAATAAGAAATAAAGAATATATCTGTAGCTATTTTGTAAACTAAGAATGTTTTAAAAA 897
Qy 1081 AACTTCTGAAATCTTAATAATTAATTAATCTGTGAGAAATC-GAATAAATATCTCTTTGTTTATT 1139
Db 898 TATTTTATTTGTAATAAATAAATTCCTTTGTTCTCTCAGAAATAAAAAAATTTTTTTTTT 957
Qy 1140 ATCGTATTCATTCGTAATAATAATAATTTCAATTTTGTAAAAA 1181
Db 958 TTGTATTAATTTTTTTTTTTTATTTATTTATTTATTTAAATA 999

RESULT 10
```

```
US-09-974-879-115
; Sequence 115, Application US/0974879
; Publication No. US20030028003A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P2
; CURRENT APPLICATION NUMBER: US/09/974,879
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239,893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/818,683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305,736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064,911
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,912
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,983
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,900
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,984
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,985
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/066,094
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,100
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,089
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,095
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,090
; PRIOR FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 115
; * LENGTH: 1797
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-974-879-115

Query Match 3.6%; Score 42.6; DB 9; Length 1797;
Best Local Similarity 61.1%; Pred. No. 1.8;
Matches 69; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 1080 AAACCTTCGAATCTTAATAATCTGAGAAATCGAACTAAATATCTCTTTGTTTATT 1139
Db 1634 AATTTCTTCTCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1693

Qy 1140 ATCGTATTCATTCGTAATAATAATTTTCATTTTGTAAAAAATTTTAAAAA 1192
Db 1694 AACTAGTACTACATATAATAATAATTTATTATTCGCTAAAAAATTTTAAAAA 1746

RESULT 11
US-09-736-116
; Sequence 116, Application US/09305736
; Publication No. US20030088078A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
```

```
FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/305,736
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: PCT/US98/23435
; EARLIER FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/064,911
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,912
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,983
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,900
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,988
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,987
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,908
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,984
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,985
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/066,094
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,100
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,089
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,095
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,090
; EARLIER FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 116
; * LENGTH: 1797
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-736-116

Query Match 3.6%; Score 42.6; DB 9; Length 1797;
Best Local Similarity 61.1%; Pred. No. 1.8;
Matches 69; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 1080 AAACCTTCGAATCTTAATAATCTGAGAAATCGAACTAAATATCTCTTTGTTTATT 1139
Db 1634 AATTTCTTCTCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1693

Qy 1140 ATCGTATTCATTCGTAATAATAATTTTCATTTTGTAAAAAATTTTAAAAA 1192
Db 1694 AACTAGTACTACATATAATAATAATTTATTATTCGCTAAAAAATTTTAAAAA 1746

RESULT 12
US-09-969-373-1172/c
; Sequence 1172, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.
; APPLICANT: Haugle, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 1172
```

```

; LENGTH: 428
; TYPE: DNA
; ORGANISM: Glycine max
; US-09-969-373-1172

Query Match      3.6%; Score 42.4; DB 10; Length 428;
Best Local Similarity 55.4%; Pred. No. 0.95; 66; Indels 0; Gaps 0;
Matches 89; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1043 AAAAGCATTATTATTCGTGAAATAAGAAATCTTACATGAAACTCTCGAAATCTTAAATAAT 1102
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 306 AAAATAATTTTAACTAATGCATAAATAATATTTATTTAAGTTGTATAAATTTATTTT 247
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1103 ACTGTGAAATCGAACTAAATATCTCTTTGTTTATTCGTATTCATTCGTATAAATA 1162
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 246 TATATAATATATATTTAAAGATATACATTTTAAATATACAATGCTATATGAATAAAT 187
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1163 ATTTTCATTTTGTAAAAAATAAAAAA 1190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 186 ATTGATATGTTATCAACAATAAAAAA 159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-10-239-676-159/c
; Sequence 159, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239, 676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 159
; LENGTH: 9515
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
; US-10-239-676-159

Query Match      3.5%; Score 42.2; DB 9; Length 9515;
Best Local Similarity 53.3%; Pred. No. 5.5;
Matches 89; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1025 AAATAATGAATCAGAGTAAAGCATTTTATTCGTGAAATAAGAATCTTACATGAACT 1084
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5360 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1085 TCTGAATCTTAATAATTAATCTGTGAGAAATCGAACTAAATATCTTTGTTTATTCGT 1144
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5300 AAACATTCCTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1145 ATTTCATTCGTAATAAATAAATTTTCATTTTGTAAAAAATAAAAAA 1191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5240 AATCTTCCCTAACCAACTTTTCTTAAAAAATTTAAAAAATAAATAAATAAATAA 5194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
US-09-864-761-15424/c
; Sequence 15424, Application US/09864761

```

```

; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeoimica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15424
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL023239.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; US-09-864-761-15424

Query Match      3.5%; Score 41.6; DB 10; Length 576;
Best Local Similarity 51.6%; Pred. No. 1.8;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 66 TTTCGAGAATGTTTCGGATCCAGATGAGCAAGAGAGAGAGATTCGAAATGTA 125
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 460 TTTCAGAAAAATCTTTTAATCCAAAAGACAAAGAGAGATGAAACATACATGAAATATG 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 126 CAAGAGATATAGAGAGATGTTTCTTCACTTCCTCGAATGATTATTGGGGGATACCAT 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 400 AAGCTCGGAGGGAACGTGCACAAAGATCTTCACTATATAGTATTAGTACATATGAC 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 186 GAGGTGTACAAGGGGATTTTGGCAAATGGGATATCTTGTACCTGGTATCATGCTTTTCCA 245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 17:21:37 ; Search time 1670 Seconds
(without alignments)
11559.892 Million cell updates/sec

Title: US-09-854-122-15
Perfect score: 1192
Sequence: 1 acgcggggaataactggaat.....gttaaaaaaaaaaaaaa 1192

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estnu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_estc.*
- 9: gb_estl.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: gb_gss.*
- 18: em_gss_hum.*
- 19: em_gss_inv.*
- 20: em_gss_pln.*
- 21: em_gss_vrt.*
- 22: em_gss_fun.*
- 23: em_gss_mam.*
- 24: em_gss_mus.*
- 25: em_gss_other.*
- 26: em_gss_pro.*
- 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	111.4	9.3	484	10	BE361455
2	97	8.1	448	13	BM500946
3	94	7.9	811	17	BH733391
4	92.6	7.8	554	10	AW830416
5	92.2	7.7	448	10	BE515611
6	91.6	7.7	828	17	BH705401

7	88.8	7.4	580	14	BQ624210
8	88.4	7.4	640	12	BF649726
9	85.8	7.2	605	10	AV558517
10	85.8	7.2	612	10	AV439900
11	84.8	7.1	361	17	BH580029
12	84.8	7.1	582	9	AI998561
13	84.6	7.1	746	12	BG586828
14	84.2	7.1	536	9	AI999125
15	84	7.0	643	10	AV441090
16	83.6	7.0	424	17	BH810739
17	83	7.0	663	10	AW309820
18	82.4	6.9	426	17	BH213349
19	82	6.9	555	14	BQ869424
20	81.4	6.8	594	10	AV439817
21	81.4	6.8	600	10	AV540478
22	81.2	6.8	377	14	LA7897
23	81.2	6.8	623	9	AI727371
24	80.6	6.8	455	10	AV564494
25	79.4	6.7	580	12	BE942160
26	79	6.6	593	14	BQ788882
27	78.6	6.6	514	12	BG300129
28	78.6	6.6	950	12	BF065941
29	77.8	6.5	531	10	AV526610
30	77.8	6.5	589	10	AV538751
31	77.8	6.5	593	10	AV539794
32	77.8	6.5	615	10	AV520358
33	77.8	6.5	685	10	AV783807
34	77.4	6.5	435	14	T88040
35	77.2	6.5	487	9	AL826245
36	77.2	6.5	572	10	AW922307
37	76.8	6.4	497	10	AW036115
38	76.4	6.4	619	10	AV439692
39	75.6	6.3	725	17	BH549521
40	75.4	6.3	385	9	AU229405
41	75.4	6.3	668	12	BF649900
42	74.8	6.3	411	10	AV520947
43	74.8	6.3	449	12	BG628738
44	74.8	6.3	516	13	BJ284052
45	74.8	6.3	556	10	AW676685

ALIGNMENTS

RESULT 1
BE361455
LOCUS DGI_172_D10_gl_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
DEFINITION BE361455 484 bp mRNA linear EST 20-JUL-2000
ACCESSION BE361455
VERSION BE361455.1 GI:9303082
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 484)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix

High quality sequence start: 32
High quality sequence stop: 476
POLYA=No.

FEATURES
source
1. .484
Location/Qualifiers

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DG1)"
/notes="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."
111 a 119 c 173 g 81 t

BASE COUNT
ORIGIN

Query Match 9.3%; Score 111.4; DB 10; Length 484;
Best Local Similarity 62.3%; Pred. No. 3.8e-13;
Matches 175; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
737 TGCATTGTTGAGTTACGAAAGAAATGTTAGAGAACCCAGTTGAAATGTGAAGAAGCTAG 796
186 TGCTGTTCTCAGTACGAGACATGCTGGAGATCGGTGGCGAGCTCAAGAAGCTGG 245
797 CTGAGTTTCATGGATGTTGTTTTCACAGCATGAGGAGAAACAAAGGATTTGTTGATGAGA 856
246 CAGCGTTTCATGGGTGGGCTTCTCTGAGGAGGACGAGGAGGATGGGGTGGTGATCAGA 305
857 TAGTTAACTTTGAGTTTCGCAATCTGAAGATCAACAGGTCAACAAAACGGATCAA 916
306 TCGTGGAGCTGTGAGTTTCGAGATCTCAAGAACGAGGACGTCACAAAGACGGGAGTA 365
917 GCTACAAATTCGAAATCGCAACAAAGCATTTCTTCAGGAAAGGTGAGGTGAGAGATTGGG 976
366 CTACGGCGTGGCATCAAGATGAGACATTTCTCAGGAGGGCGAGTCTGGCGACTGGA 425
977 CAAACTATCTAACCTCGGAATGATTAAGAAATCGGAGACG 1017
426 AAAAATCATGACTGTGGACATGGCGCGAGGCTGGATAAG 466

RESULT 2
BM500946 448 bp mRNA linear EST 14-FEB-2002
DEFINITION PAC00000000793 Pioneer AF-1 array Zea mays cDNA, mRNA sequence.
ACCESSION BM500946
VERSION BM500946.1 GI:186660595
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.

REFERENCE 1 (bases 1 to 448)
AUTHORS Hunter, B.G., Beatty, M., Singletary, G., Hamaker, B., Larkins, B.A. and Jung, R.
TITLE Maize opaque endosperm mutations create extensive changes in patterns of gene expression
JOURNAL Unpublished (2002)
COMMENT Contact: Jung R
Pioneer Hi-Bred International, Inc.
7300 NW 62nd Ave., P.O. Box 1004, Johnston, IA 50131-1004, USA
Tel: 515 270 5934
Fax: 515 254 2619
Email: rudolf.jung@pioneer.com.

FEATURES
source
1. .448
Location/Qualifiers
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="Pioneer AF-1 array"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI"
113 a 86 c 154 g 94 t 1 others

Query Match 7.9%; Score 94; DB 17; Length 811;
Best Local Similarity 52.9%; Pred. No. 1.4e-09;
Matches 277; Conservative 0; Mismatches 235; Indels 12; Gaps 3;
478 CCGAGGTGTTTCGCGGACACATCCCATCTCTTTGTCGCGGCTCTGTTTGAATCG 537
132 CCGAGACATTCGCAACCCACGTCCTTCGTTGCTTCAGGACTCCATCGAGGACCA 191

Query Match 8.1%; Score 97; DB 13; Length 448;
Best Local Similarity 60.9%; Pred. No. 3.8e-10;
Matches 176; Conservative 0; Mismatches 110; Indels 3; Gaps 1;
737 TGCATTGTTGAGTTACGAAAGAAATGTTAGAGAACCCAGTTGAAATGTGAAGAAGCTAG 796
11 TGTGTTCTCCTCGGTACGAGGAGATGCTGATTGATCCCGAGGCACACGTCAGGAAGCTCG 70
797 CTGAGTTTCATGGATGTTGTTTTCACAGCATGAGGAGAAACAAAGGATTTGTTGATGAGA 856
71 CCAAGTTTCATGGGTGTTGATTTCTGAGGAGAGAGGAGCAGCGGGTGGTGAGCGCCA 130
857 TAGTTAACTTTGAGTTTCGCAATCTGAAGATCAACAGGTCAACAAAACGGATCAA 916
131 TCGTGGAGCTGTGAGTCTGAGCTGGCAAGATGAGGACATGAGGATGAACAGAAATGG---AA 187
917 GCTACAAATTCGAAATCGCAACACAGCATTTCTTCAGGAAAGGTGAGGTGAGAGATTGGG 976
188 GCAATATGTTGGGGTCAAGAACGAAAGCTACTTTCAGGAAAGGAGTTGCTGGGACTGGA 247
977 CAAACTATCTAACCTCGGAAATGATTAAAGAAATCGAGACGCGCGGAAA 1025
248 GCAACCATATGACCGGACATGGCGACAGCTAGACAAAGTCTGTAGA 296

RESULT 3

BH733391 811 bp DNA linear GSS 20-FEB-2002
BOMIU60TR BO_2_3_KB Brassica oleracea genomic clone BOMIU60, DNA
sequence.
ACCESSION BH733391
VERSION BH733391.1 GI:18838786
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea

REFERENCE 1 (bases 1 to 811)
AUTHORS Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOMIU60TF
Contact: Chris Town
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source
1. .811
Location/Qualifiers
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOMIU60"

/clone_lib="BO_2_3_KB"
/notes="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
222 a 184 c 202 g 203 t

Query Match 7.9%; Score 94; DB 17; Length 811;
Best Local Similarity 52.9%; Pred. No. 1.4e-09;
Matches 277; Conservative 0; Mismatches 235; Indels 12; Gaps 3;
478 CCGAGGTGTTTCGCGGACACATCCCATCTCTTTGTCGCGGCTCTGTTTGAATCG 537
132 CCGAGACATTCGCAACCCACGTCCTTCGTTGCTTCAGGACTCCATCGAGGACCA 191


```

QY 526 GTTTGAATCGGGAACAAATAATCAATATATAAGCGCAACGTAAGAGTACATTTGTG 585
Db 317 CTGACACACTTTCCTTGCAGAGATTGTGTATGTGTAGGAACGTGAAGGACACATTTGATC 376
QY 586 TCTTTTGGAAATTCGGCA------TCTGATTACCCCGCAAGCTATTGGACCTCGA 638
Db 377 TCAAGTTGGTTTTTTTGGCCATTCCTTTCTTAAATACAAACCAAGCAATCTCGA 436
QY 639 AAAGACGTTGATATCTTCGATCGGGAATCTCTTTTGTGGACCGGAATGGAATTTCCA 698
Db 437 GTCTATGTTTCAAGAGTTCTGCAATGGAATCTATCTATTGACCTTTTGGGAATATCT 496
QY 699 AGCGGAGTTACCA-ATGCGGCTCTACTAATTAATCAAACTTG--CTATTGTTAGTTACGA 755
Db 497 CTTGGGCTACTGAGAGAAGAGCTTGAAGAGACCCCAAGCATGTCCTTTTCATGAGGTAGGA 556
QY 756 AGAAATGTTAGAGAAGCCAGTTCAAAATGTGAAGAGCTAGCTGACTTCATGGGATGTGG 815
Db 557 GGAATGAAACAGAGCCAGCTGATCAGATCAAGAGACTAGCCGACTTCTTGGGTTGTCC 616
QY 816 GTTCACAGACGATGAGGAGAAACAAAGGATTGTTGATGAGATAGTTAAACTTTGTAGCTT 875
Db 617 TTTTACTAGCAGAGAAGATATATGGAGTTGTAGACAAGGCTTGTGGACCTTTGCTCTCT 676
QY 876 CGCAATCTGAAGAAATCAACAGGTGAACAAA 907
Db 677 GCCTAATCTGAGCAGTTTGGAGGTTAACAAA 708

RESULT 7
BQ624210
LOCUS
DEFINITION
sinensis cDNA clone USDA-FP_01301 5', mRNA sequence.
ACCESSION
BQ624210
VERSION
sinensis
SOURCE
Citrus sinensis
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE
1 (bases 1 to 580)
Bauscher M., McKendree W., Dang P., Chaparro J., Shatters R., Hunter
W. and Niedz R.
Expressed sequence tags isolated from entire sweet orange (C.
sinensis L. Osbeck) seedling
Unpublished (2003)
Contact: Michael Bauscher
US Horticultural Research
USDA - ARS
2001 South Rock Rd., Fort Pierce, FL 34945, USA
Tel: (772) 462-5918
Fax: (772) 462-5961
Email: mbausher@ushrl.ars.usda.gov
Seq primer: T3 Primer.
Location/Qualifiers
1..580
/organism="Citrus sinensis"
/cultivar="Ridge Pineapple"
/db_xref="taxon:2711"
/clone="USDA-FP_01301"
/clone_lib="Ridge pineapple sweet orange entire seedling"
/tissue_type="entire seedling"
/dev_stage="50 days after germination"
/lab_host="XJ1-Blue"
/notes="vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; A high quality EST with at least 200 contiguous
bases at Trace Tuner score of 20 or better"
171 a 110 c 139 g 160 t

BASE COUNT
ORIGIN

```

```

Query Match 7.4%; Score 88.8; DB 14; Length 580;
Best Local Similarity 49.2%; Pred. No. 1.8e-08;
Matches 262; Conservative 0; Mismatches 267; Indels 3; Gaps 1;

QY 487 TTTTGGCGGACATCCCATACTCTTTGTTGCGGGCTGTGTTTGAATTCGGGAACAAA 546
Db 1 TTTTGAACGCATGTCGGCATGCTTTACTACAGGTTCCATCTGAAATCTGGCTGCGA 60
QY 547 ATCATCAATATTAAGCGCAACCGTAAGAGTACATTTGTCTGTTTGGAAATTTGGCAAT 606
Db 61 AATGTGTAGTTTGTAGGAACCCCTGGACCAATTCATCTCAGAGTGGCTATTATAGCT 120
QY 607 CTGATTAAACCCCGCAAGTATTGACCTCGAAAGAGCGTTGATATCTTCGATCGGA 666
Db 121 AGAACTCAGGACAAGGAGCCATCTGACCTAGCGGAAGCTTTTGGAGGGCCCTGTAAATGA 180
QY 667 ATCTCTCTTTTGGACCGGAATTTCCAAAGCGGAGTTCCACCAATCGGGGCTCTACT 726
Db 181 ATCCAGATTTTGGACCCCATTTGGGAACATGGTGGTGTACTGGAGAGCCAGTATAGAA 240
QY 727 AATTC---AAACTTGTCTATTGTTGAGTTACGAAGAAATGTTAGAGAAGCCAGTTGAAAAT 783
Db 241 CAACCGGATAAGATATTCTTCTTGAATATGAAGATCTCAAGAAGATATCGCCTCTTGC 300
QY 784 GTGAAGAAGCTAGCTGAGTTTCATGGGATGTGGTTCACAGAGTATGAGAGAAACAAAGG 843
Db 301 ATTAACAGATTGGCGGATTTCTTGGGATGTCCTCTTTTCGGAAGAGGAGTACCCCAAGGT 360
QY 844 ATTGTTGATGAGATAGTTAAACTTTGTAGCTTCGCAATCTGAAGATCAACAGAGTGAAC 903
Db 361 GTGGTGAAGAAATCTCAAGCTATGATGTTTGGATATATCCAAACTTTGGAAGTCAAC 420
QY 904 AAAAAAGGATCAAGCTACAAATTCGAAATTCGAAATTCGAAATTCCTTCAGGAAAGGTGAG 963
Db 421 AAAACGGTAGAGCGTACCAATGGGTGGAATAATCTCACTACTTGGAGAAAGGTGAA 480
QY 964 GTGAGAGATTGGCAAACTATCTAAGCTCGGAAATGATTAAGAAACTGAGA 1015
Db 481 GTTGGAGACTGGAAAAATTTATTAACCCATCGATGTCAGAGCGTTTGGAGA 532

RESULT 8
BQ649726
LOCUS
DEFINITION
clone NF082E05C1F1037 Elicited call culture Medicago truncatula cDNA
ACCESSION
BQ649726
VERSION
BQ649726.1
SOURCE
barrel medic.
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
1 (bases 1 to 640)
Torres-Jerez I., Scott A.D., Harris A.R., Gonzales R.A., Bell C.J.,
Flores H.R., Inman J.T., Weller J.W. and May G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
Center for Medicago Genomics Research
Unpublished (2000)
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 640 Std Error: 0.00
Plate: 082 row: E column: 05
Seq primer: TCACAGGAAACAGCTATGAC.
Location/Qualifiers
1..640
source

```

```

/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF082E05EC"
/clone_lib="Elicited cell culture"
/tissue_type="Cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14
days. Cells were induced six days after subculture"
/notes="Vector: Lambda Zap; Cells were induced with yeast
cell wall extracts equivalent to 50ug/ml glucose in the
final concentration. Samples were taken at 0.5, 1, 12 and
24 hours after induction. Equal amounts of RNA from each
time point were pooled and used for mRNA isolation."
BASE COUNT      208 a      86 g      210 t
ORIGIN
Query Match      7.4%; Score 88.4; DB 12; Length 640;
Best Local Similarity 49.6%; Pred. No. 2.1e-08;
Matches 255; Conservative 0; Mismatches 256; Indels 3; Gaps 1;
QY 104 AGGAAGAAGATTCCAAATGTACAGAGATATAGAGAGATTGTTCTTCACTTCCCTCGA 163
DB 112 AAGACAAAGTCACTTAGCCNAGAAACACAGAACTAATCTTCTTCCAGAGAGA 171
QY 164 ATGATTATGGGGGATACCATGAGTTGTACAGGATTTTGGCAATGGGATATCTTG 223
DB 172 AAGGTTCGAGAACACCTTATATTTATCTATTCACAGATTTTGTGCGCAACGCTGAA 231
QY 224 TACCTGTATCATGGCTTCGAGATATTTCAAGGCTCGAGAGACGACATATCTCTTA 283
DB 232 TCCAGGCCATAACACATTTCCAAAGCATTTCCAAAGCTAAAGAAAGTGATGTTTGTG 291
QY 284 CGACTCTTCCAAAGGCTGGAAACGACATGGAGGACCTGACGTTTGCCATCCTAACAC 343
DB 292 CAACCGTACCAATACGAGACACTTGGTTAAAGCTCTTACCTATGCCATATGATC 351
QY 344 GAGATGTTAACCCATCATACCGACACATCCATTTTGTGTTTCAACCCCTCATPGT 403
DB 352 GCCAAACCATTTTCATTTTCATCCAAAACCATCTCTTGTAGTTTCAATCCACATGC 411
QY 404 GTGTTCAAAATTTGGAGTATTTGTACATGGTGAAGAAATACGATCCGAGCTCGATA 463
DB 412 TTGTTCTCTTCATTTGATGATACAGTTTATGATGATGACAAATTCCTGATTTGCTA 471
QY 464 TGTGAATGAATGCCGAGGTTGTTGCCGAGACATCCCATACTCTTGTGTCGGCGGT 523
DB 472 AAATTCATGA--GCCTAGACTTTTGGTACACATATCTCTTGTACTCATTTGCCAAT 528
QY 524 CTGTTTGAATCGGGAAACAAATATCAATATAAGCCGCAACCGTAAGAGTACATTTG 583
DB 529 CAATCAAGGTTCCAAATGCAAAATAGTTTATATTTGTAGGAACCCCTTTTGATACTTCA 588
QY 584 TGCTCTTTTGGAAATTTGGCAATCTGATTAAACC 617
DB 589 TCTCTTCTGGACTTTTGGCAACAAATTAAGCC 622

RESULT 9
AV558517/c
LOCUS
DEFINITION
AV558517 Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone SQ101b10F 3', mRNA sequence.
ACCESSION
AV558517
VERSION
AV558517.1 GI:8729943
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 605)
REFERENCE
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..605
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ101b10F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      199 a      149 c      80 g      177 t
ORIGIN
Query Match      7.2%; Score 85.8; DB 10; Length 605;
Best Local Similarity 59.1%; Pred. No. 7.4e-08;
Matches 166; Conservative 0; Mismatches 112; Indels 3; Gaps 1;
QY 733 AACTTCCTATTGTTGAGTTACGAAGAAATGTTAGAGAAGCCAGTTCGAAATGTGAAGAAG 792
DB 471 ATGTTCTTTTCTATGAAGTACGAAGAGATAATTGAGGAGCTCTCGTGTTCAGTCAAGAGA 412
QY 793 CTAGCTGAGTTCATGGGATGTGGTTTCACAGACGATGAGGAGAAACAAGGGATTTGTTGAT 852
DB 411 CTCGCCGAGTCTTGGAAATGTCATTCCACAGGAAGAGAGAGTGGATCGGTGGAG 352
QY 853 GAGATAGTTAACTTTGTAGCTTCGACAAATCTGAAGAATCAACAGGTGAACAAACCGGA 912
DB 351 GAGATCTTGAAGTTGTGTAGTTTACGAAATTTAAGCAATTTTGAGGTTTAAAGAAATGGG 292
QY 913 TCAAGCTACAAATTCGAAATTCGACAAACAGCATTTCTTCAGGAAAGGTGAGGTGAGAGAT 972
DB 291 ACNACAGAAATTT---GGTGTAGATTCCTCAGGTGTTCTTTAGGAAGGTGAAGTTGTTGAT 235
QY 973 TGGGCAAACTACTATAACGTCGGAATGATTAAGAAACTGGA 1013
DB 234 TGAAGAATCATCTTACGCCACAATGGCGAAAACCTTTGA 194

RESULT 10
AV439900/c
LOCUS
DEFINITION
AV439900 Arabidopsis thaliana above-ground organ two to six-week
old Arabidopsis thaliana cDNA clone APD20g12_f_3', mRNA sequence.
ACCESSION
AV439900
VERSION
AV439900.1 GI:7610249
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 612)
REFERENCE
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..612

```

```
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db xref="taxon:3702"
/clone_lib="Arabidopsis thaliana above-ground organ two to
six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      200 a 152 c 84 g 176 t
ORIGIN

Query Match      7.2%; Score 85.8; DB 10; Length 612;
Best Local Similarity 59.1%; Pred. No. 7.4e-08;
Matches 166; Conservative 0; Mismatches 112; Indels 3; Gaps 1;

QY 733 AACTTCCTATTGTTGAGTTCAGGAAGAAATGTTAGAGAGCCAGTTCGAAATGTGAGAG 792
Db 447 AATGTTCTTTTCATGAAGTACGAAGAGATAATGAGAGGCTTCGTGTTCAAGTCAAGAGA 388
QY 793 CTAGCTCAGTTCATGGGATGTGGTTTCACAGACGATCAGGAGAAACAAGGGATTGTTGAT 852
Db 387 CTCGCCGAGTTCCTGGAATGTCATTCCACCAAGAGAGAGAAAGTGGATCGGTGGAG 328
QY 853 GAGATAGTTAAACTTTGTAGTTCGCAATCTCAAGAAATCAACAGGTGAACAAAAACGGA 912
Db 327 GAGATCTGAAGTTGTGTAGTTACGAAATTTAAGCAATTTGAGAGTTTAATAGAAATGGG 268
QY 913 TCAGCTACATTCGAAATTCGACACAGCAATTTCTTCAGGAAAGGTGAGGTGAGAGAT 972
Db 267 ACAACAGGAAT---GGTGTAGATTTCTCAGGTGTTCTTTAGGAAAGGTGAAGTTGGTGAT 211
QY 973 TGGGCAAACTATCTAAGTCGGAATGATTAAAGAACTGGA 1013
Db 210 TGAAGAAATCACTTACGCACAAATGGCGAAACCTTTGA 170

RESULT 11
LOCUS BH580029/c 361 bp DNA linear GSS 14-DEC-2001
DEFINITION BOGLB46TF BOGL Brassica oleracea genomic clone BOGLB46, DNA
sequence.
ACCESSION BH580029
VERSION BH580029.1 GI:17832310
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
REFERENCE Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
AUTHORS Whole genome shotgun sequencing of Brassica oleracea
TITLE Unpublished (2001)
JOURNAL Other GSSs: BOGLB46TR
COMMENT Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .361
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db xref="taxon:3712"
/clone="BOGLB46"
/clone_lib="BOGL"
/notes="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
```

```
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      93 a 111 c 44 g 113 t
ORIGIN

Query Match      7.1%; Score 84.8; DB 17; Length 361;
Best Local Similarity 59.6%; Pred. No. 1.4e-07;
Matches 162; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

QY 740 TATTGTGAGTTACGAAGAATAATGTTAGAGAAGCCAGTTCGAAATGTGAAGAGCTAGCTG 799
Db 341 TGTTCCTTAAGTACGAAGATCTGAAGAAGATATTGAGACTAACTTGAAGAAGCTTGCAA 282
QY 800 AGTTCAATGGATGTGGGTTTCACAGACGATGAGGAGAAACAAGGATGTTGATGAGATAG 859
Db 281 GTTTCCTGGGGTTCCTTTTCAGCGAAGAAGAGAGAAAGGCTGTTGTAAGGCGGTAG 222
QY 860 TTAACCTTTGTAGCTTCGACAAATCTGAAGAATCAACAGGTGACAAACGATCAAGCT 919
Db 221 CGATCTGTGTAGTTCGAGAGTCTAAAGAAGTTGGAGGTGAACAAGTTCGAACAATCGA 162
QY 920 ACAATTCGAAATTCGACAAACAGCATTTCTTCAGGAAAGGTGAGGTGAGAGATTGGGCAA 979
Db 161 TCAA---GACTTGAAGATAGTACTTGTTTAGGAAAGAGAGGTTAGTGGCTGGAGAA 105
QY 980 ACTATCTAAGTCGGAATGATTAAAGAACTG 1011
Db 104 ACTATTGTGCGCTGTACAAGTGGAGAGATTG 73

RESULT 12
LOCUS AI998561/c 582 bp mRNA linear EST 08-SEP-1999
DEFINITION thaliana cDNA clone 701546181, mRNA sequence.
ACCESSION AI998561
VERSION AI998561.1 GI:5845466
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 582)
Chen,J., Moniyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,
Wang,X., Hillman,J., Guebler,K., Kim,C., Doyle,M., Brzoska,P.,
Gorgone,G., Burns,D., Griffin,J., Mouganoutou,M., Nguyen,D., Tan,R.,
Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,
Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobrigha,A., Murry,L.,
Turner,C., Krikorian,S., Elder,L. and Hanson,D.
Arabidopsis thaliana Gene Expression MicroArray
Unpublished (1999)
Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.
Location/Qualifiers
1. .582
/organism="Arabidopsis thaliana"
/cultivar="Columbia Col-0"
/db xref="taxon:3702"
/clone="701546181"
/clone_lib="A. thaliana, Columbia Col-0, rosette-2"
/tissue_type="rosette"
/dev_stage="4 - 7 weeks"
/notes="Vector: pSPORT; Site 1: NotI; Site 2: SalI; CDNA
library was derived from untreated rosette tissue from
Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks.
Plants were grown in 1:1 peat moss/vermiculite/perlite
soil at 22 deg. C +/- 3 deg. C under constant light, and
watered with fertilizer. cDNA synthesis was initiated
```

using a NotI-oligo(dT) primer. Double-stranded cDNA was blunted, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

BASE COUNT 174 a 133 c 81 g 186 t 8 others

ORIGIN

Query Match 7.1%; Score 84.6; DB 9; Length 582;
Best Local Similarity 58.7%; Pred. No. 1.2e-07;
Matches 165; Conservative 0; Mismatches 113; Indels 3; Gaps 1;

QY 737 TGCATTTGTTGAGTTACGAAATGTTAGAGAGCCAGTTGAAATGTGAAAGCTAG 796
|||||
Db 395 TGCATTTTATGAAGTTGAAGAGATGAAGACAGAACCTCGTACCAGATCAAGAAATTTG 336
|||||

QY 797 CTGAGTTTCATGGGATGGGTTCCACAGACCATGAGGAGAAACAAGGGATTGTTGATGAGA 856
|||||

Db 335 CCAGTTCTTAGTGTGTCCTTACTAAGGAAGAAGAGAGCGGATCGGTGGATGAGA 276
|||||

QY 857 TAGTTAACTTTGTAGCTTCGCAATCTGAAGATCAACAGGTGAACAAAAGCGATCAA 916
|||||

Db 275 TTATCGATCTTTGTTCTACGTAACTGAGCAGTTTGGAGATCAATAAGACCGG---AA 219
|||||

QY 917 GCTACAAATTCGAAATCGACAACAGCATTTCTTCAGGAAGGTGAGGTGAGATGGG 976
|||||

Db 218 AATTGAATTCGTGAGAGAAACAAAATGTTTTCCGTAAGGAGAAAGTTGGTGATTGGA 159
|||||

QY 977 CAACTATCTAACTGCGGAATGATTAAAGAACTGGAGACG 1017
|||||

Db 158 AGAATATTGATCTCTGGAATGGAGAAATAATCGACATG 118
|||||

RESULT 13
BG586828

LOCUS

DEFINITION

EST488597 MRAM Medicago truncatula/Glomus versiforme mixed EST
Library cDNA clone pHEM-44L16 5' end, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

ESTs from roots of Medicago truncatula after colonization with Glomus versiforme, 2001
Unpublished (2001)
Contact: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name: N385895e TIGR sequence name: MTDCX68TK More
Information is available at: <http://www.medicago.org>
Seq primer: SKmod (CTA GAA CTA 959 GAT CC).

FEATURES

source

1. 746
/organisms="Medicago truncatula/Glomus versiforme mixed EST
library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="pHEM-44L16"
/tissue="roots"
/dev_stage="Roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The library was
made from a mixture of RNA from each of these stages."
/lab_hosts="E. coli strain XL0LR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 238 a 146 c 119 g 243 t

ORIGIN

Query Match 7.1%; Score 84.6; DB 12; Length 746;
Best Local Similarity 49.1%; Pred. No. 1.2e-07;
Matches 253; Conservative 0; Mismatches 259; Indels 3; Gaps 1;

QY 104 AGGAACAAGATCCAAATGTACAAGATATAGAGAGATGTTCTTCACATTCCTCGA 163
|||||

Db 116 AGAACAAGGTCACTTACCGAAGAAACAAGAACTAATCTCTCTTCCAAAGAGA 175
|||||

QY 164 ATGATTATTGGGGGATACCATGAGGTTTGTACAAGGATTTTGGCAATGGGATATCTTG 223
|||||

Db 176 AGGTTGGAGACACCTTATATTATCTATTCCAAGATTTTGGTCCCAACAGCTGAA 235
|||||

QY 224 TACCTGGTATCATGGCTTCGAAGATAATTTCAAGGCTCGAGAGACGGACATTTATCCTTA 283
|||||

Db 236 TCCAAGCCATAACCACTTTCCAAAGCAATTTCCAAGCTAAAGAAAGTGATGTTTGTG 295
|||||

QY 284 CGACTCTTCCAAGGCTGGAACGACATGACGAGGACCTGAGTTGGCATCCTCAACAC 343
|||||

Db 296 CAACCGTACCAAAATCAGGACAACTTGGTTAAAGCTCTTACCTATGCCATTTATGAATC 355
|||||

QY 344 GAGATGTTAAACCAACCCATCATCACGACATCCACTTTTGTCTTCAACCCCTCATTCGT 403
|||||

Db 356 GCCAAACCACTTCATTTTCATCCAAACCACTTTGCTTAGTTTCAATCCACATGATC 415
|||||

QY 404 GTGTTCAAAATTTGGAGTATTTGTACATGGGTAGAGAAAATACGATGCGAGACCTCGATA 463
|||||

Db 416 TTGTTCTCTTTCATTTGAGTATACAGTTTATGTTAAGCATGACACAAATTCCTGATTGCTA 475
|||||

QY 464 TGTGTAATGAATCGCGAGGTTGTTGCGGACACATCCCATCTCTTGTGTTCCGCGCT 523
|||||

Db 476 AATTCATGA---GCCTAGACTTTTGGTACACATATTCCTTTGACTCATGTCCAAAT 532
|||||

QY 524 CTGTTTGAATCGGGAACAAAATCATCAATATAAGCCGCAACCGTAAGAGTACATTTG 583
|||||

Db 533 CAATCAAGGTTCCAAATTCAAAATAGTTTATATTTGTAGGAACCCCTTTTGATACTTCA 592
|||||

QY 584 TGTCTTTTGGAAATTTGGCAATCTGATTAACCCC 618
|||||

Db 593 TCTCTTCTTGGACTTTTGGCAACAATTAAGCCAC 627
|||||

RESULT 14
-A1999125/c

LOCUS

DEFINITION

thaliana cDNA clone 701554544, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 536)
Chen, J., Himiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
Wang, X., Hallman, J., Guebler, K., Kim, C., Doyle, M., Brzoska, P.,
Gorgone, G., Burns, D., Griffin, J., Mouanoutou, M., Nguyen, D., Tan, R.,
Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T.,
Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobriga, A., Murry, L.,
Turner, C., Krikorian, S., Elder, D. and Hanson, D.
Arabidopsis thaliana Gene Expression MicroArray

TITLE

JOURNAL
COMMENT

Unpublished (1999)
Contact: David Smoller, Ph.D.
Genom Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

FEATURES

source

1. .536
/organism="Arabidopsis thaliana"
/cultivar="Columbia Col-0"
/db_xref="taxon:3702"
/clone="70155454"
/clone_lib="A. thaliana, Columbia Col-0, rosette-3"
/tissue_type="rosette"
/dev_stage="4 - 7 weeks"
/notes="Vector: pSPORT; Site 1: NotI; Site 2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunt-ended, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

BASE COUNT 154 a 147 c 84 g 150 t 1 others
ORIGIN

Query Match 7.1%; Score 84.2; DB 9; Length 536;
Best Local Similarity 60.2%; Pred. No. 1.6e-07;
Matches 162; Conservative 0; Mismatches 98; Indels 9; Gaps 1;
QY 740 TATTGTTGAGTACGAGAAATGTTAGAGAGCCAGTTGAAATGTCGAAGAGCTAGCTG 799
DB 448 TGTTCTCAAGTACGAGACGATGAGCTGATCTCTTTACCGTACGTGAAGAGTCTGGCTG 389
QY 800 AGTTTCATGGGATGTGGGTTTCACAGACGATGAGGAGAAACAAGGATTTGTTGATGAGATAG 859
DB 388 AGTTTATGGTCTAGTTTACAGCCGAGGAGGAGAAAGGTTGTTGAGAAAGTGG 329
QY 860 TTAACCTTTTGTAGTTCGCAATCTGAAGATC-----AACAGGTGAACAAAAACG 910
DB 328 TGAATCTTTTCAGCTTCGAGACGTTGAAGAATCTTGAAGCTTAACAAAGGGGAGAGACA 269
QY 911 GATCAGACTACATTCGAAATCGACACAGCAAGCATTTCTTCAGGAAGAGTGAAGTGAAG 970
DB 268 GAGAGAGATCGTCTGGTGTGTTTACCGGAATAGCGCGTATTTTCAGGAAGGAAAGGTGGGAG 209
QY 971 ATTGGGCAAACTATCTAAACGTCCGAAATG 999
DB 208 ATTGTCGAACATCTGACTCCGGAGATG 180

RESULT 15

AV441090/c

LOCUS

AV441090 643 bp mRNA linear EST 14-NOV-2000
DEFINITION Arabidopsis thaliana above-ground organ two to six-week
old Arabidopsis thaliana cDNA clone AP227a12_f_3', mRNA sequence.

ACCESSION

AV441090

VERSION

AV441090.1

KEYWORDS

EST.

SOURCE

ORGANISM

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 643)

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation

of 12,028 non-redundant expressed sequence tags from normalized and

size-selected cDNA libraries

JOURNAL

MEDLINE

COMMENT

DNA Res. 7, 175-180 (2000)

20363093

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source

1. .643
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AP227a12 f"
/clone_lib="Arabidopsis thaliana above-ground organ two to six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 208 a 152 c 94 g 189 t
ORIGIN

Query Match 7.0%; Score 84; DB 10; Length 643;

Best Local Similarity 60.4%; Pred. No. 1.7e-07;

Matches 157; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

QY 733 AACTTGTCTATTGTTGAGTTACGAAGAAATGTTAGAGAAGCCAGTTGAAATGTGAAGAAG 792

DB 581 AAAGTCTTATTTTAAAGTACGAGGATCTCAAAGAGACATCGAGACCACCTTGAAGAAG 522

QY 793 CTAGCTGAGTTCATGGGATGTGGGTTTCACAGACGATGAGGAGAAACAAGGATTTGTTGAT 852

DB 521 CTAGCAAGTTTCTTAGGACTTCCTTTCCACCGAAGACAGGAAACAAAAGGGAGTTGTGAAA 462

QY 853 GAGATAGTTAACTTTTGTAGCTTCGACATCTGAAGATCAACAGGTGAACAAAAACGGA 912

DB 461 GCTATCGCTGATCTGTGTAGCTTTGAGAATCTGAAGAAGTTGGAGGTGAAC---AAGTCA 405

QY 913 TCAAGCTACAAATTCGAAATCGACAACAACCATTTCTTCAGGAAAGGTGAGGTGAGAGAT 972

DB 404 AGCAATTTGATCCAGAACTATGAGAACCCGTTCTTTTAGGAAAGGAGAGTGAAGTAT 345

QY 973 TGGGCAAACTATCTAAACGTC 992

DB 344 TTGTTAACTATTGTGCGC 325

Search completed: June 10, 2003, 18:50:17

Job time : 1674 secs

THIS PAGE BLANK (USPTO)